

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
10 January 2002 (10.01.2002)

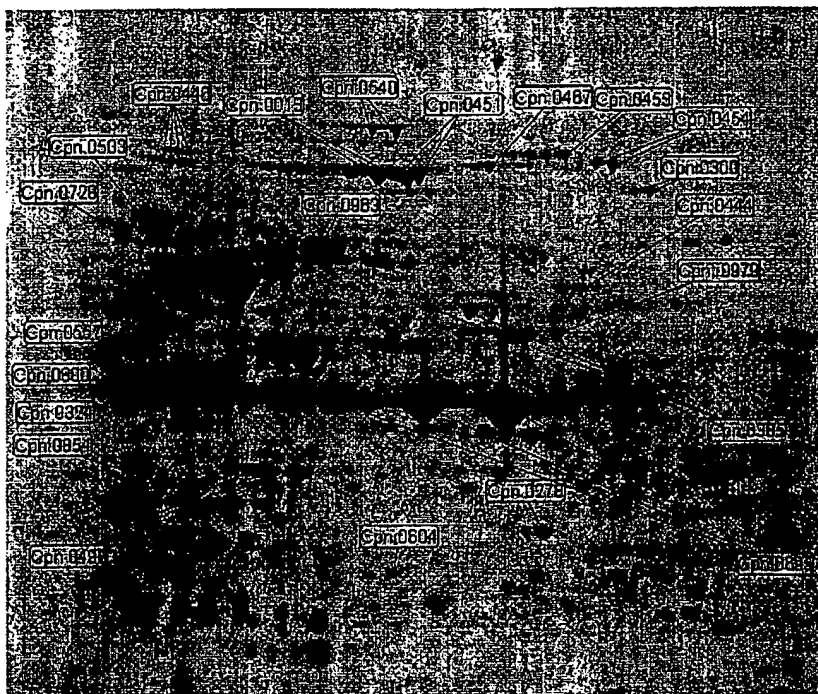
PCT

(10) International Publication Number  
**WO 02/02606 A2**

- (51) International Patent Classification<sup>7</sup>: **C07K 14/295**,  
C12N 15/31, A61K 39/118
- (21) International Application Number: **PCT/IB01/01445**
- (22) International Filing Date: **3 July 2001 (03.07.2001)**
- (25) Filing Language: **English**
- (26) Publication Language: **English**
- (30) Priority Data:
- |           |                                |    |
|-----------|--------------------------------|----|
| 0016363.4 | 3 July 2000 (03.07.2000)       | GB |
| 0017047.2 | 11 July 2000 (11.07.2000)      | GB |
| 0017983.8 | 21 July 2000 (21.07.2000)      | GB |
| 0019368.0 | 7 August 2000 (07.08.2000)     | GB |
| 0020440.4 | 18 August 2000 (18.08.2000)    | GB |
| 0022583.9 | 14 September 2000 (14.09.2000) | GB |
| 0027549.5 | 10 November 2000 (10.11.2000)  | GB |
| 0031706.5 | 22 December 2000 (22.12.2000)  | GB |
- (71) Applicant (for all designated States except US): **CHIRON S.P.A.** [IT/IT]; Via Fiorentina, 1, I-53100 Siena (IT).
- (72) Inventors; and
- (73) Inventors/Applicants (for US only): **RATTI, Giulio** [IT/IT]; Chiron S.p.A., Via Fiorentina, 1, I-53100 Siena (IT). **GRANDI, Guido** [IT/IT]; Chiron S.p.A., Via Fiorentina, 1, I-53100 Siena (IT).
- (74) Agents: **HALLYBONE, Huw, George et al.**; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European

[Continued on next page]

(54) Title: **IMMUNISATION AGAINST *CHLAMYDIA PNEUMONIAE***



(57) Abstract: The published genomic of *Chlamydia pneumoniae* reveals over 1000 putative encoded proteins but does not itself indicate which of these might be useful antigens for immunisation and vaccination or for diagnosis. This difficulty is addressed by the invention, which provides a number of *C. pneumoniae* protein sequences suitable for vaccine production and development and/or for diagnostic purposes.



patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

**Published:**

- *without international search report and to be republished upon receipt of that report*

## IMMUNISATION AGAINST *CHLAMYDIA PNEUMONIAE*

All documents cited herein are incorporated by reference in their entirety.

### TECHNICAL FIELD

This invention is in the field of immunisation against chlamydial infection, in particular against infection by *Chlamydia pneumoniae*.

### BACKGROUND ART

*Chlamydiae* are obligate intracellular parasites of eukaryotic cells which are responsible for endemic sexually transmitted infections and various other disease syndromes. They occupy an exclusive eubacterial phylogenic branch, having no close relationship to any other known organisms – they are classified in their own order (*Chlamydiales*) which contains a single family (*Chlamydiaceae*) which in turn contains a single genus (*Chlamydia*). A particular characteristic of the *Chlamydiae* is their unique life cycle, in which the bacterium alternates between two morphologically distinct forms: an extracellular infective form (elementary bodies, EB) and an intracellular non-infective form (reticulate bodies, RB). The life cycle is completed with the re-organization of RB into EB, which subsequently leave the disrupted host cell ready to infect further cells.

Four chlamydial species are currently known – *C.trachomatis*, *C.pneumoniae*, *C.pecorum* and *C.psittaci* [e.g. Raulston (1995) *Mol Microbiol* 15:607-616; Everett (2000) *Vet Microbiol* 75:109-126]. *C.pneumoniae* is closely related to *C.trachomatis*, as the whole genome comparison of at least two isolates from each species has shown [Kalman *et al.* (1999) *Nature Genetics* 21:385-389; Read *et al.* (2000) *Nucleic Acids Res* 28:1397-406; Stephens *et al.* (1998) *Science* 282:754-759]. Based on surface reaction with patient immune sera, the current view is that only one serotype of *C.pneumoniae* exists world-wide.

*C.pneumoniae* is a common cause of human respiratory disease. It was first isolated from the conjunctiva of a child in Taiwan in 1965, and was established as a major respiratory pathogen in 1983. In the USA, *C.pneumoniae* causes approximately 10% of community-acquired pneumonia and 5% of pharyngitis, bronchitis, and sinusitis.

More recently, the spectrum of *C.pneumoniae* infections has been extended to include atherosclerosis, coronary heart disease, carotid artery stenosis, myocardial infarction, cerebrovascular disease, aortic aneurysm, claudication, and stroke. The association of *C.pneumoniae* with atherosclerosis is corroborated by the presence of the organism in atherosclerotic lesions throughout the arterial tree and the near absence of the organism in healthy arterial tissue. *C.pneumoniae* has also been isolated from coronary and carotid atheromatous plaques. The bacterium has also been associated with other acute and chronic respiratory diseases (e.g. otitis media, chronic obstructive pulmonary disease, pulmonary exacerbation of cystic fibrosis) as a result of sero-epidemiologic observations, case reports, isolation or direct detection of the organism in specimens, and successful

response to anti-chlamydial antibiotics. To determine whether chronic infection plays a role in initiation or progression of disease, intervention studies in humans have been initiated, and animal models of *C.pneumoniae* infection have been developed.

Considerable knowledge of the epidemiology of *C.pneumoniae* infection has been derived from serologic studies using the *C.pneumoniae*-specific microimmunofluorescence test. Infection is ubiquitous, and it is estimated that virtually everyone is infected at some point in life, with common re-infection. Antibodies against *C.pneumoniae* are rare in children under the age of 5, except in developing and tropical countries. Antibody prevalence increases rapidly at ages 5 to 14, reaching 50% at the age of 20, and continuing to increase slowly to ~80% by age 70.

A current hypothesis is that *C.pneumoniae* can persist in an asymptomatic low-grade infection in very large sections of the human population. When this condition occurs, it is believed that the presence of *C.pneumoniae*, and/or the effects of the host reaction to the bacterium, can cause or help progress of cardiovascular illness.

It is not yet clear whether *C.pneumoniae* is actually a causative agent of cardiovascular disease, or whether it is just artefactually associated with it. It has been shown, however, that *C.pneumoniae* infection can induce LDL oxidation by human monocytes [Kalayoglu *et al.* (1999) *J. Infect. Dis.* 180:780-90; Kalayoglu *et al.* (1999) *Am. Heart J.* 138:S488-490]. As LDL oxidation products are highly atherogenic, this observation provides a possible mechanism whereby *C.pneumoniae* may cause atheromatous degeneration. If a causative effect is confirmed, vaccination (prophylactic and therapeutic) will be universally recommended.

Genomic sequence information has been published for *C.pneumoniae* [Kalman *et al.* (1999) *supra*; Read *et al.* (2000) *supra*; Shirai *et al.* (2000) *J. Infect. Dis.* 181(Suppl 3):S524-S527; WO99/27105; WO00/27994] and is available from GenBank. Sequencing efforts have not, however, focused on vaccination, and the availability of genomic sequence does not in itself indicate which of the >1000 genes might encode useful antigens for immunisation and vaccination. WO99/27105, for instance, implies that every one of the 1296 ORFs identified in the *C.pneumoniae* strain CM1 genome is a useful vaccine antigen.

It is thus an object of the present invention to identify antigens useful for vaccine production and development from amongst the many proteins present in *C.pneumoniae*. It is a further object to identify antigens useful for diagnosis (*e.g.* immunodiagnosis) of *C.pneumoniae*.

## DISCLOSURE OF THE INVENTION

The invention provides proteins comprising the *C.pneumoniae* amino acid sequences disclosed in the examples.

It also provides proteins comprising sequences which share at least *x*% sequence identity with the *C.pneumoniae* amino acid sequences disclosed in the examples. Depending on the particular



sequence,  $x$  is preferably 50% or more (e.g. 60%, 70%, 80%, 90%, 95%, 99% or more). These include mutants and allelic variants. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH  
5 program (Oxford Molecular), using an affine gap search with parameters *gap open penalty*=12 and *gap extension penalty*=1.

The invention further provides proteins comprising fragments of the *C.pneumoniae* amino acid sequences disclosed in the examples. The fragments should comprise at least  $n$  consecutive amino acids from the sequences and, depending on the particular sequence,  $n$  is 7 or more (e.g. 8, 10, 12,  
10 14, 16, 18, 20, 30, 40, 50, 75, 100 or more). Preferably the fragments comprise one or more epitope(s) from the sequence. Other preferred fragments omit a signal peptide.

The proteins of the invention can, of course, be prepared by various means (e.g. native expression, recombinant expression, purification from cell culture, chemical synthesis *etc.*) and in various forms (e.g. native, fusions *etc.*). They are preferably prepared in substantially pure form (*ie.* substantially  
15 free from other *C.pneumoniae* or host cell proteins). Heterologous expression in *E.coli* is a preferred preparative route.

According to a further aspect, the invention provides nucleic acid comprising the *C.pneumoniae* nucleotide sequences disclosed in the examples. In addition, the invention provides nucleic acid comprising sequences which share at least  $x\%$  sequence identity with the *C.pneumoniae* nucleotide  
20 sequences disclosed in the examples. Depending on the particular sequence,  $x$  is preferably 50% or more (e.g. 60%, 70%, 80%, 90%, 95%, 99% or more).

Furthermore, the invention provides nucleic acid which can hybridise to the *C.pneumoniae* nucleic acid disclosed in the examples, preferably under "high stringency" conditions (e.g. 65°C in a 0.1xSSC, 0.5% SDS solution).

25 Nucleic acid comprising fragments of these sequences are also provided. These should comprise at least  $n$  consecutive nucleotides from the *C.pneumoniae* sequences and, depending on the particular sequence,  $n$  is 10 or more (e.g. 12, 14, 15, 18, 20, 25, 30, 35, 40, 50, 75, 100, 200, 300 or more).

According to a further aspect, the invention provides nucleic acid encoding the proteins and protein fragments of the invention.

30 It should also be appreciated that the invention provides nucleic acid comprising sequences complementary to those described above (e.g. for antisense or probing purposes).

Nucleic acid according to the invention can, of course, be prepared in many ways (e.g. by chemical synthesis, from genomic or cDNA libraries, from the organism itself *etc.*) and can take various forms (e.g. single stranded, double stranded, vectors, probes *etc.*).

In addition, the term "nucleic acid" includes DNA and RNA, and also their analogues, such as those containing modified backbones, and also peptide nucleic acids (PNA) *etc.*

According to a further aspect, the invention provides vectors comprising nucleotide sequences of the invention (*e.g.* cloning or expression vectors) and host cells transformed therewith.

- 5 According to a further aspect, the invention provides immunogenic compositions comprising protein and/or nucleic acid according to the invention. These compositions are suitable for immunisation and vaccination purposes. Vaccines of the invention may be prophylactic or therapeutic, and will typically comprise an antigen which can induce antibodies capable of inhibiting (a) chlamydial adhesion, (b) chlamydial entry, and/or (c) successful replication within the host cell. The vaccines  
10 preferably induce any cell-mediated T-cell responses which are necessary for chlamydial clearance from the host.

- The invention also provides nucleic acid or protein according to the invention for use as medicaments (*e.g.* as vaccines). It also provides the use of nucleic acid or protein according to the invention in the manufacture of a medicament (*e.g.* a vaccine or an immunogenic composition) for  
15 treating or preventing infection due to *C.pneumoniae*.

The invention also provides a method of treating (*e.g.* immunising) a patient, comprising administering to the patient a therapeutically effective amount of nucleic acid or protein according to the invention.

According to further aspects, the invention provides various processes.

- 20 A process for producing proteins of the invention is provided, comprising the step of culturing a host cell according to the invention under conditions which induce protein expression.

A process for producing protein or nucleic acid of the invention is provided, wherein the protein or nucleic acid is synthesised in part or in whole using chemical means.

- A process for detecting *C.pneumoniae* in a sample is provided, wherein the sample is contacted with  
25 an antibody which binds to a protein of the invention.

A summary of standard techniques and procedures which may be employed in order to perform the invention (*e.g.* to utilise the disclosed sequences for immunisation) follows. This summary is not a limitation on the invention but, rather, gives examples that may be used, but are not required.

#### General

- 30 The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature *e.g.* Sambrook *Molecular Cloning; A Laboratory Manual, Second Edition* (1989) and *Third Edition* (2001); *DNA Cloning, Volumes I and ii* (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); *Transcription and Translation* (B.D. Hames & S.J. Higgins eds. 1984); *Animal Cell Culture* (R.I.  
35

- 5 Freshney ed. 1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide to Molecular Cloning* (1984); the *Methods in Enzymology* series (Academic Press, Inc.), especially volumes 154 & 155; *Gene Transfer Vectors for Mammalian Cells* (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Mayer and Walker, eds. (1987), *Immunochemical Methods in Cell and Molecular Biology* (Academic Press, London); Scopes, (1987) *Protein Purification: Principles and Practice*, Second Edition (Springer-Verlag, N.Y.), and *Handbook of Experimental Immunology, Volumes I-IV* (D.M. Weir and C. C. Blackwell eds 1986).

Standard abbreviations for nucleotides and amino acids are used in this specification.

#### Definitions

- 10 A composition containing X is "substantially free of" Y when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95% or even 99% by weight.

The term "comprising" means "including" as well as "consisting" e.g. a composition "comprising" X may consist exclusively of X or may include something additional to X, such as X+Y.

- 15 The term "heterologous" refers to two biological components that are not found together in nature. The components may be host cells, genes, or regulatory regions, such as promoters. Although the heterologous components are not found together in nature, they can function together, as when a promoter heterologous to a gene is operably linked to the gene. Another example is where a Chlamydial sequence is heterologous to a mouse host cell. A further examples would be two epitopes from the same or different proteins which have been  
20 assembled in a single protein in an arrangement not found in nature.

- An "origin of replication" is a polynucleotide sequence that initiates and regulates replication of polynucleotides, such as an expression vector. The origin of replication behaves as an autonomous unit of polynucleotide replication within a cell, capable of replication under its own control. An origin of replication may be needed for a vector to replicate in a particular host cell. With certain origins of replication, an expression vector can be  
25 reproduced at a high copy number in the presence of the appropriate proteins within the cell. Examples of origins are the autonomously replicating sequences, which are effective in yeast; and the viral T-antigen, effective in COS-7 cells.

- A "mutant" sequence is defined as DNA, RNA or amino acid sequence differing from but having sequence identity with the native or disclosed sequence. Depending on the particular sequence, the degree of sequence  
30 identity between the native or disclosed sequence and the mutant sequence is preferably greater than 50% (e.g. 60%, 70%, 80%, 90%, 95%, 99% or more, calculated using the Smith-Waterman algorithm as described above). As used herein, an "allelic variant" of a nucleic acid molecule, or region, for which nucleic acid sequence is provided herein is a nucleic acid molecule, or region, that occurs essentially at the same locus in the genome of another or second isolate, and that, due to natural variation caused by, for example, mutation or recombination,  
35 has a similar but not identical nucleic acid sequence. A coding region allelic variant typically encodes a protein having similar activity to that of the protein encoded by the gene to which it is being compared. An allelic variant can also comprise an alteration in the 5' or 3' untranslated regions of the gene, such as in regulatory control regions (e.g. see US patent 5,753,235).

### Expression systems

The Chlamydial nucleotide sequences can be expressed in a variety of different expression systems; for example those used with mammalian cells, baculoviruses, plants, bacteria, and yeast.

#### i. Mammalian Systems

5 Mammalian expression systems are known in the art. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*e.g.* structural gene) into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, usually located 25-30 base pairs (bp) upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA  
10 synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element, usually located within 100 to 200 bp upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation [Sambrook et al. (1989) "Expression of Cloned Genes in Mammalian Cells." In *Molecular Cloning: A Laboratory Manual*, 2nd ed.].

Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding  
15 mammalian viral genes provide particularly useful promoter sequences. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter (Ad MLP), and herpes simplex virus promoter. In addition, sequences derived from non-viral genes, such as the murine metallothionein gene, also provide useful promoter sequences. Expression may be either constitutive or regulated (inducible), depending on the promoter can be induced with glucocorticoid in hormone-responsive  
20 cells.

The presence of an enhancer element (enhancer), combined with the promoter elements described above, will usually increase expression levels. An enhancer is a regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to homologous or heterologous promoters, with synthesis beginning at the normal RNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription  
25 initiation site, in either normal or flipped orientation, or at a distance of more than 1000 nucleotides from the promoter [Maniatis et al. (1987) *Science* 236:1237; Alberts et al. (1989) *Molecular Biology of the Cell*, 2nd ed.]. Enhancer elements derived from viruses may be particularly useful, because they usually have a broader host range. Examples include the SV40 early gene enhancer [Dijkema et al (1985) *EMBO J.* 4:761] and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus [Gorman et al.  
30 (1982) *PNAS USA* 79:6777] and from human cytomegalovirus [Boshart et al. (1985) *Cell* 41:521]. Additionally, some enhancers are regulatable and become active only in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) *Trends Genet.* 2:215; Maniatis et al. (1987) *Science* 236:1237].

A DNA molecule may be expressed intracellularly in mammalian cells. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein  
35 will always be a methionine, which is encoded by the ATG start codon. If desired, the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in mammalian cells. Preferably, there are processing sites encoded between the leader

fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The adenovirus tripartite leader is an example of a leader sequence that provides for secretion of a foreign protein in mammalian cells.

- 5 Usually, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-transcriptional cleavage and polyadenylation [Birnstiel et al. (1985) *Cell* 41:349; Proudfoot and Whitelaw (1988) "Termination and 3' end processing of eukaryotic RNA. In *Transcription and splicing* (ed. B.D. Hames and D.M. Glover); Proudfoot (1989) *Trends Biochem. Sci.* 14:105]. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator/polyadenylation signals include those derived from SV40 [Sambrook et al (1989) "Expression of cloned genes in cultured mammalian cells." In *Molecular Cloning: A Laboratory Manual*].

- 15 Usually, the above described components, comprising a promoter, polyadenylation signal, and transcription termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (*e.g.* plasmids) capable of stable maintenance in a host, such as mammalian cells or bacteria. Mammalian replication systems include those derived from animal viruses, which require trans-acting factors to replicate. For example, plasmids containing the replication systems of papovaviruses, such as SV40 [Gluzman (1981) *Cell* 23:175] or polyomavirus, replicate to extremely high copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replicaton systems, thus allowing it to be maintained, for example, in mammalian cells for expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-bacteria shuttle vectors include pMT2 [Kaufman et al. (1989) *Mol. Cell. Biol.* 9:946] and pHEBO [Shimizu et al. (1986) *Mol. Cell. Biol.* 6:1074].

- 25 The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene-mediated transfection, protoplast fusion, electroporation, encapsulation of polynucleotide(s) in liposomes, direct microinjection of the DNA into nuclei.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (*e.g.* Hep G2), and a number of other cell lines.

## 35 ii. Baculovirus Systems

- The polynucleotide encoding the protein can also be inserted into a suitable insect expression vector, and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence
- 40

homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.

After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit). These techniques are generally known to those skilled in the art and fully described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) (hereinafter "Summers and Smith").

Prior to inserting the DNA sequence encoding the protein into the baculovirus genome, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are usually assembled into an intermediate transplacement construct (transfer vector). This construct may contain a single gene and operably linked regulatory elements; multiple genes, each, with its own set of operably linked regulatory elements; or multiple genes, regulated by the same set of regulatory elements. Intermediate transplacement constructs are often maintained in a replicon, such as an extrachromosomal element (*e.g.* plasmids) capable of stable maintenance in a host, such as a bacterium. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification.

Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; see Luckow and Summers, *Virology* (1989) 17:31.

The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) *Ann. Rev. Microbiol.*, 42:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of replication for selection and propagation in *E. coli*.

Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (*e.g.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression may be either regulated or constitutive.

Structural genes, abundantly transcribed at late times in a viral infection cycle, provide particularly useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedron protein, Friesen et al., (1986) "The Regulation of Baculovirus Gene Expression," in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler); EPO Publ. Nos. 127 839 and 155 476; and the gene encoding the p10 protein, Vlak et al., (1988), *J. Gen. Virol.* 69:765.

DNA encoding suitable signal sequences can be derived from genes for secreted insect or baculovirus proteins, such as the baculovirus polyhedrin gene (Carbonell et al. (1988) *Gene*, 73:409). Alternatively, since the signals

for mammalian cell posttranslational modifications (such as signal peptide cleavage, proteolytic cleavage, and phosphorylation) appear to be recognized by insect cells, and the signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells, leaders of non-insect origin, such as those derived from genes encoding human  $\alpha$ -interferon, Maeda et al., (1985), *Nature* 315:592; human gastrin-releasing peptide, Lebacqz-Verheyden et al., (1988), *Molec. Cell. Biol.* 8:3129; human IL-2, Smith et al., (1985) *Proc. Nat'l Acad. Sci. USA*, 82:8404; mouse IL-3, (Miyajima et al., (1987) *Gene* 58:273; and human glucocerebrosidase, Martin et al. (1988) *DNA*, 7:99, can also be used to provide for secretion in insects.

A recombinant polypeptide or polyprotein may be expressed intracellularly or, if it is expressed with the proper regulatory sequences, it can be secreted. Good intracellular expression of nonfused foreign proteins usually requires heterologous genes that ideally have a short leader sequence containing suitable translation initiation signals preceding an ATG start signal. If desired, methionine at the N-terminus may be cleaved from the mature protein by *in vitro* incubation with cyanogen bromide.

Alternatively, recombinant polyproteins or proteins which are not naturally secreted can be secreted from the insect cell by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in insects. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the translocation of the protein into the endoplasmic reticulum.

After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer vector and the genomic DNA of wild type baculovirus -- usually by co-transfection. The promoter and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summers and Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) 3:2156; and Luckow and Summers (1989)). For example, the insertion can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays* 4:91. The DNA sequence, when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.

The newly formed baculovirus expression vector is subsequently packaged into an infectious recombinant baculovirus. Homologous recombination occurs at low frequency (between ~1% and ~5%); thus, the majority of the virus produced after cotransfection is still wild-type virus. Therefore, a method is necessary to identify recombinant viruses. An advantage of the expression system is a visual screen allowing recombinant viruses to be distinguished. The polyhedrin protein, which is produced by the native virus, is produced at very high levels in the nuclei of infected cells at late times after viral infection. Accumulated polyhedrin protein forms occlusion bodies that also contain embedded particles. These occlusion bodies, up to 15 $\mu$ m in size, are highly refractile, giving them a bright shiny appearance that is readily visualized under the light microscope. Cells infected with recombinant viruses lack occlusion bodies. To distinguish recombinant virus from wild-type virus, the transfection supernatant is plaqued onto a monolayer of insect cells by techniques known to those skilled in the art. Namely, the plaques are screened under the light microscope for the presence (indicative of wild-type virus)

or absence (indicative of recombinant virus) of occlusion bodies. "Current Protocols in Microbiology" Vol. 2 (Ausubel et al. eds) at 16.8 (Supp. 10, 1990); Summers & Smith, *supra*; Miller et al. (1989).

Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*: *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni* (WO 89/046699; Carbonell et al., (1985) *J. Virol.* 56:153; Wright (1986) *Nature* 321:718; Smith et al., (1983) *Mol. Cell. Biol.* 3:2156; and see generally, Fraser, et al. (1989) *In Vitro Cell. Dev. Biol.* 25:225).

Cells and cell culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression system; cell culture technology is generally known to those skilled in the art. See, e.g. Summers and Smith *supra*.

The modified insect cells may then be grown in an appropriate nutrient medium, which allows for stable maintenance of the plasmid(s) present in the modified insect host. Where the expression product gene is under inducible control, the host may be grown to high density, and expression induced. Alternatively, where expression is constitutive, the product will be continuously expressed into the medium and the nutrient medium must be continuously circulated, while removing the product of interest and augmenting depleted nutrients. The product may be purified by such techniques as chromatography, e.g. HPLC, affinity chromatography, ion exchange chromatography, etc.; electrophoresis; density gradient centrifugation; solvent extraction, or the like. As appropriate, the product may be further purified, as required, so as to remove substantially any insect proteins which are also secreted in the medium or result from lysis of insect cells, so as to provide a product which is at least substantially free of host debris, e.g. proteins, lipids and polysaccharides.

In order to obtain protein expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant protein encoding sequence. These conditions will vary, dependent upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

### iii. Plant Systems

There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as: US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry* 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in Vaulcombe et al., *Mol. Gen. Genet.* 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 3:407-418 (1984); Rogers, *J. Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et al., *Nucleic Acids Research* 15:2515-2535 (1987); Wirsal et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by gibberellic acid can be found in R.L. Jones and J. MacMillin, *Gibberellins*: in: *Advanced Plant Physiology*, Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52. References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987)



Typically, using techniques known in the art, a desired polynucleotide sequence is inserted into an expression cassette comprising genetic regulatory elements designed for operation in plants. The expression cassette is inserted into a desired expression vector with companion sequences upstream and downstream from the expression cassette suitable for expression in a plant host. The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the vector to permit the vectors to move DNA from an original cloning host, such as bacteria, to the desired plant host. The basic bacterial/plant vector construct will preferably provide a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for *Agrobacterium* transformations, T DNA sequences for *Agrobacterium*-mediated transfer to plant chromosomes. Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a plant cell has been transformed. A general review of suitable markers, for example for the members of the grass family, is found in Wilink and Dons, 1993, *Plant Mol. Biol. Rept.*, 11(2):165-185.

Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences and the like for homologous recombination as well as Ti sequences which permit random insertion of a heterologous expression cassette into a plant genome. Suitable prokaryote selectable markers include resistance toward antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art.

The nucleic acid molecules of the subject invention may be included into an expression cassette for expression of the protein(s) of interest. Usually, there will be only one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous protein encoding sequence the following elements, a promoter region, plant 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

A heterologous coding sequence may be for any protein relating to the present invention. The sequence encoding the protein of interest will encode a signal peptide which allows processing and translocation of the protein, as appropriate, and will usually lack any sequence which might result in the binding of the desired protein of the invention to a membrane. Since, for the most part, the transcriptional initiation region will be for a gene which is expressed and translocated during germination, by employing the signal peptide which provides for translocation, one may also provide for translocation of the protein of interest. In this way, the protein(s) of interest will be translocated from the cells in which they are expressed and may be efficiently harvested. Typically secretion in seeds are across the aleurone or scutellar epithelium layer into the endosperm of the seed. While it is not required that the protein be secreted from the cells in which the protein is produced, this facilitates the isolation and purification of the recombinant protein.

Since the ultimate expression of the desired gene product will be in a eucaryotic cell it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed out as introns by the host's splicosome machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code, Reed and Maniatis, *Cell* 41:95-105, 1985.

The vector can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, *Mol. Gen. Genet.*, 202:179-185, 1985. The genetic material may also be

transferred into the plant cell by using polyethylene glycol, Krens, et al., *Nature*, 296, 72-74, 1982. Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface, Klein, et al., *Nature*, 327, 70-73, 1987 and Knudsen and Muller, 1991, *Planta*, 185:330-336 teaching particle bombardment of barley endosperm to create transgenic barley. Yet another method of introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley, et al., *Proc. Natl. Acad. Sci. USA*, 79, 1859-1863, 1982.

The vector may also be introduced into the plant cells by electroporation. (Fromm et al., *Proc. Natl Acad. Sci. USA* 82:5824, 1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersion*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Hererocallis*, *Nemesia*, *Pelargonium*, *Panicum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browaalia*, *Glycine*, *Lolium*, *Zea*, *Triticum*, *Sorghum*, and *Datura*.

Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted. Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

In some plant cell culture systems, the desired protein of the invention may be excreted or alternatively, the protein may be extracted from the whole plant. Where the desired protein of the invention is secreted into the medium, it may be collected. Alternatively, the embryos and embryoless-half seeds or other plant tissue may be mechanically disrupted to release any secreted protein between cells and tissues. The mixture may be suspended in a buffer solution to retrieve soluble proteins. Conventional protein isolation and purification methods will be then used to purify the recombinant protein. Parameters of time, temperature pH, oxygen, and volumes will be adjusted through routine methods to optimize expression and recovery of heterologous protein.

#### iv. Bacterial Systems

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (e.g. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in *Escherichia coli* (*E. coli*) [Raibaud *et al.* (1984) *Annu. Rev. Genet.* 18:173]. Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (*lac*) [Chang *et al.* (1977) *Nature* 198:1056], and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (*trp*) [Goeddel *et al.* (1980) *Nuc. Acids Res.* 8:4057; Yelverton *et al.* (1981) *Nucl. Acids Res.* 9:731; US patent 4,738,921; EP-A-0036776 and EP-A-0121775]. The *g-lactamase* (*bla*) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In *Interferon* 3 (ed. I. Gresser)], bacteriophage lambda PL [Shimatake *et al.* (1981) *Nature* 292:128] and T5 [US patent 4,689,406] promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [US patent 4,551,433]. For example, the *tac* promoter is a hybrid *trp-lac* promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor [Amann *et al.* (1983) *Gene* 25:167; de Boer *et al.* (1983) *Proc. Natl. Acad. Sci.* 80:21]. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier *et al.* (1986) *J. Mol. Biol.* 189:113; Tabor *et al.* (1985) *Proc Natl. Acad. Sci.* 82:1074]. In addition, a hybrid promoter can also be comprised of a bacteriophage promoter and an *E. coli* operator region (EPO-A-0 267 851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In *E. coli*, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine *et al.* (1975) *Nature* 254:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' end of *E. coli* 16S rRNA [Steitz *et al.* (1979) "Genetic signals and nucleotide sequences in messenger RNA." In *Biological*

*Regulation and Development: Gene Expression* (ed. R.F. Goldberger)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site [Sambrook *et al.* (1989) "Expression of cloned genes in *Escherichia coli*." In *Molecular Cloning: A Laboratory Manual*].

5 A DNA molecule may be expressed intracellularly. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide or by either *in vivo* or *in vitro* incubation with a bacterial methionine N-terminal peptidase (EPO-A-0 219 237).

10 Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene [Nagai *et al.* (1984) *Nature* 309:810]. Fusion proteins can also be  
15 made with sequences from the *lacZ* [Jia *et al.* (1987) *Gene* 60:197], *trpE* [Allen *et al.* (1987) *J. Biotechnol.* 5:93; Makoff *et al.* (1989) *J. Gen. Microbiol.* 135:11], and *Chey* [EP-A-0 324 647] genes. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*e.g.* ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign  
20 protein. Through this method, native foreign protein can be isolated [Miller *et al.* (1989) *Bio/Technology* 7:698].

Alternatively, foreign proteins can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the foreign protein in bacteria [US patent 4,336,336]. The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is  
25 either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the *E. coli* outer membrane protein gene (*ompA*) [Masui *et al.* (1983), in: *Experimental Manipulation of Gene Expression*; Ghayeb *et al.* (1984) *EMBO J.* 3:2437] and the *E. coli* alkaline phosphatase signal sequence (*phoA*) [Oka *et al.* (1985) *Proc. Natl. Acad. Sci.* 82:7212]. As an additional example, the signal sequence of the alpha-amylase gene from various *Bacillus* strains can be used to secrete heterologous proteins from *B. subtilis* [Palva  
30 *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 244 042].

Usually, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the  
35 translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in terminating transcription. Examples include transcription termination sequences derived from genes with strong promoters, such as the *trp* gene in *E. coli* as well as other biosynthetic genes.

Usually, the above described components, comprising a promoter, signal sequence (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (e.g. plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a prokaryotic host either for expression or for cloning and amplification. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host.

Alternatively, the expression constructs can be integrated into the bacterial genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to the bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bacterial chromosome. For example, integrating vectors constructed with DNA from various *Bacillus* strains integrate into the *Bacillus* chromosome (EP-A- 0 127 328). Integrating vectors may also be comprised of bacteriophage or transposon sequences.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and may include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline [Davies *et al.* (1978) *Annu. Rev. Microbiol.* 32:469]. Selectable markers may also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are usually comprised of a selectable market that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, *inter alia*, the following bacteria: *Bacillus subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541], *Escherichia coli* [Shimatake *et al.* (1981) *Nature* 292:128; Amann *et al.* (1985) *Gene* 40:183; Studier *et al.* (1986) *J. Mol. Biol.* 189:113; EP-A-0 036 776, EP-A-0 136 829 and EP-A-0 136 907], *Streptococcus cremoris* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655]; *Streptococcus lividans* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655], *Streptomyces lividans* [US patent 4,745,056].

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with  $\text{CaCl}_2$  or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. See e.g. [Masson *et al.* (1989) *FEMS Microbiol. Lett.* 60:273; Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, *Bacillus*], [Miller *et al.* (1988) *Proc. Natl. Acad. Sci.* 85:856; Wang *et al.* (1990) *J. Bacteriol.* 172:949, *Campylobacter*], [Cohen *et al.* (1973) *Proc. Natl. Acad. Sci.* 69:2110; Dower *et al.* (1988) *Nucleic Acids Res.* 16:6127; Kushner (1978) "An improved method for transformation of *Escherichia coli* with ColE1-derived plasmids. In *Genetic*

*Engineering: Proceedings of the International Symposium on Genetic Engineering* (eds. H.W. Boyer and S. Nicosia); Mandel *et al.* (1970) *J. Mol. Biol.* 53:159; Taketo (1988) *Biochim. Biophys. Acta* 949:318; Escherichia], [Chassy *et al.* (1987) *FEMS Microbiol. Lett.* 44:173 Lactobacillus]; [Fiedler *et al.* (1988) *Anal. Biochem* 170:38, Pseudomonas]; [Augustin *et al.* (1990) *FEMS Microbiol. Lett.* 66:203, Staphylococcus],  
 5 [Barany *et al.* (1980) *J. Bacteriol.* 144:698; Harlander (1987) "Transformation of Streptococcus lactis by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss III); Perry *et al.* (1981) *Infect. Immun.* 32:1295; Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655; Somkuti *et al.* (1987) *Proc. 4th Eur. Cong. Biotechnology* 1:412, Streptococcus].

#### v. Yeast Expression

10 Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (e.g. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the "TATA Box") and a transcription initiation site. A yeast promoter may  
 15 also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the  
 20 metabolic pathway provide particularly useful promoter sequences. Examples include alcohol dehydrogenase (ADH) (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences [Myanohara *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:1].

25 In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the *ADH2*,  
 30 *GAL4*, *GAL10*, OR *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, *inter alia*, [Cohen *et al.* (1980) *Proc. Natl. Acad. Sci. USA* 77:1078; Henikoff *et al.* (1981) *Nature* 283:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol.* 96:119;  
 35 Hollenberg *et al.* (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast *Saccharomyces cerevisiae*," in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon *et al.* (1980) *Gene* 11:163; Panthier *et al.* (1980) *Curr. Genet.* 2:109;].

A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always

be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See *e.g.* EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*e.g.* ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (*e.g.* WO88/024066).

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the genes for invertase (EP-A-0012873; JPO 62,096,086) and A-factor (US patent 4,588,684). Alternatively, leaders of non-yeast origin exist, such as an interferon leader, that also provide for secretion in yeast (EP-A-0060057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008; EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alphafactor. (*e.g.* see WO 89/02463.)

Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (*e.g.* plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YEp24 [Botstein *et al.* (1979) *Gene* 8:17-24], pCI/1 [Brake *et al.* (1984) *Proc. Natl. Acad. Sci USA* 81:4642-4646], and YRp17 [Stinchcomb *et al.* (1982) *J. Mol. Biol.* 158:157]. In addition, a replicon may be either a high or low copy number plasmid. A high copy

number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Enter a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See *e.g.* Brake *et al.*, *supra*.

- 5 Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome [Orr-Weaver *et al.* (1983) *Methods in Enzymol.* 101:228-245]. An integrating vector may be  
10 directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver *et al.*, *supra*. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced [Rine *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the chromosome and flanking the  
15 expression construct in the vector, which can result in the stable integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may include biosynthetic genes that can be expressed in the yeast host, such as *ADE2*, *HIS4*, *LEU2*, *TRP1*, and *ALG7*, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable  
20 marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of *CUP1* allows yeast to grow in the presence of copper ions [Butt *et al.* (1987) *Microbiol. Rev.* 51:351].

Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or  
25 developed into an integrating vector, as described above.

Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, *inter alia*, the following yeasts: *Candida albicans* [Kurtz, *et al.* (1986) *Mol. Cell. Biol.* 6:142], *Candida maltosa* [Kunze, *et al.* (1985) *J. Basic Microbiol.* 25:141], *Hansenula polymorpha* [Gleeson, *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302], *Kluyveromyces fragilis* [Das, *et al.* (1984) *J. Bacteriol.* 158:1165], *Kluyveromyces lactis* [De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:737; Van den Berg *et al.* (1990) *Bio/Technology* 8:135], *Pichia guilliermondii* [Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141], *Pichia pastoris* [Cregg, *et al.* (1985) *Mol. Cell. Biol.* 5:3376; US Patent Nos. 4,837,148 and 4,929,555], *Saccharomyces cerevisiae* [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163], *Schizosaccharomyces pombe* [Beach and Nurse (1981) *Nature* 300:706], and *Yarrowia lipolytica* [Davidow, *et al.* (1985) *Curr. Genet.* 10:380471 Gaillardin, *et al.* (1985) *Curr. Genet.* 10:49].  
30  
35

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See *e.g.* [Kurtz *et al.* (1986) *Mol. Cell. Biol.* 6:142; Kunze  
40 *et al.* (1985) *J. Basic Microbiol.* 25:141; *Candida*]; [Gleeson *et al.* (1986) *J. Gen. Microbiol.* 132:3459;



Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302; Hansenula]; [Das *et al.* (1984) *J. Bacteriol.* 158:1165; De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:1165; Van den Berg *et al.* (1990) *Bio/Technology* 8:135; Kluyveromyces]; [Cregg *et al.* (1985) *Mol. Cell. Biol.* 5:3376; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; US Patents 4,837,148 & 4,929,555; Pichia]; [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163 *Saccharomyces*]; [Beach & Nurse (1981) *Nature* 300:706; *Schizosaccharomyces*]; [Davidow *et al.* (1985) *Curr. Genet.* 10:39; Gaillardin *et al.* (1985) *Curr. Genet.* 10:49; *Yarrowia*].

#### Pharmaceutical Compositions

Pharmaceutical compositions can comprise polypeptides and/or nucleic acid of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of either polypeptides, antibodies, or polynucleotides of the claimed invention.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation can be determined by routine experimentation and is within the judgement of the clinician.

For purposes of the present invention, an effective dose will be from about 0.01 mg/kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier.

### Delivery Methods

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

5 Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (e.g. see WO98/20734), needles, and gene guns or hypodermic sprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

### Vaccines

10 Vaccines according to the invention may either be prophylactic (i.e. to prevent infection) or therapeutic (i.e. to treat disease after infection).

Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with "pharmaceutically acceptable carriers," which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, etc. pathogens.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59<sup>TM</sup> (WO 90/14837; Chapter 10 in Vaccine design: the subunit and adjuvant approach, eds. Powell & Newman, Plenum Press 1995), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalene, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi<sup>TM</sup> adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox<sup>TM</sup>); (3) saponin adjuvants, such as Stimulon<sup>TM</sup> (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (e.g. IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, etc.), interferons (e.g. gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc; and (6) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59<sup>TM</sup> are preferred.

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-*sn*-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), *etc.*

5 The immunogenic compositions (*e.g.* the immunising antigen/immunogen/polypeptide/protein/ nucleic acid, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such as water, saline, glycerol, ethanol, *etc.* Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

10 Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.

15 Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic or immunogenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated (*e.g.* nonhuman primate, primate, *etc.*), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be  
20 determined through routine trials.

The immunogenic compositions are conventionally administered parenterally, *e.g.* by injection, either subcutaneously, intramuscularly, or transdermally/transcutaneously (*e.g.* WO98/20734). Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be  
25 administered in conjunction with other immunoregulatory agents.

As an alternative to protein-based vaccines, DNA vaccination may be employed [*e.g.* Robinson & Torres (1997) *Seminars in Immunology* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648; see later herein].

#### Gene Delivery Vehicles

30 Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic of the invention, to be delivered to the mammal for expression in the mammal, can be administered either locally or systemically. These constructs can utilize viral or non-viral vector approaches in *in vivo* or *ex vivo* modality. Expression of such coding sequence can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence *in vivo* can be either constitutive or regulated.

35 The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral, adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus, picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) *Cancer Gene Therapy* 1:51-64; Kimura (1994) *Human Gene Therapy* 5:845-852; Connelly (1995) *Human Gene Therapy* 6:185-193; and Kaplitt (1994) *Nature Genetics* 6:148-153.

Retroviral vectors are well known in the art and we contemplate that any retroviral gene therapy vector is employable in the invention, including B, C and D type retroviruses, xenotropic retroviruses (for example, NZB-X1, NZB-X2 and NZB9-1 (see O'Neill (1985) *J. Virol.* 53:160) polytropic retroviruses e.g. MCF and MCF-MLV (see Kelly (1983) *J. Virol.* 45:291), spumaviruses and lentiviruses. See RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985.

Portions of the retroviral gene therapy vector may be derived from different retroviruses. For example, retrovector LTRs may be derived from a Murine Sarcoma Virus, a tRNA binding site from a Rous Sarcoma Virus, a packaging signal from a Murine Leukemia Virus, and an origin of second strand synthesis from an Avian Leukosis Virus.

- 10 These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA by incorporation of a chimeric integrase enzyme into the retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.
- 15 Packaging cell lines suitable for use with the above-described retrovirus vectors are well known in the art, are readily prepared (see WO95/30763 and WO92/05266), and can be used to create producer cell lines (also termed vector cell lines or "VCLs") for the production of recombinant vector particles. Preferably, the packaging cell lines are made from human parent cells (e.g. HT1080 cells) or mink parent cell lines, which eliminates inactivation in human serum.
- 20 Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J Virol* 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC No. VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from
- 25 depositories or collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or isolated from known sources using commonly available techniques.

Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; WO89/05349, WO89/09271, WO90/02806, WO90/07936, WO94/03622, WO93/25698, WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53 (1993) 83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

Human adenoviral gene therapy vectors are also known in the art and employable in this invention. See, for example, Berkner (1988) *Biotechniques* 6:616 and Rosenfeld (1991) *Science* 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors employable in this invention include those described in the above referenced documents and in WO94/12649, WO93/03769, WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671,

WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654. Alternatively, administration of DNA linked to killed adenovirus as described in Curiel (1992) *Hum. Gene Ther.* 3:147-154 may be employed. The gene delivery vehicles of the invention also include  
5 adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava, WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in which the native D-sequences are modified by substitution of nucleotides, such that at least 5 native nucleotides and up to 18 native nucleotides, preferably at least 10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the  
10 remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive nucleotides in each AAV inverted terminal repeat (*ie.* there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the native D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of  
15 which are disclosed in Nahreini (1993) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in US Patent 5,478,745. Still other vectors are those disclosed in Carter US Patent 4,797,368 and Muzyczka US Patent 5,139,941, Chartejee US Patent 5,474,935, and Kotin WO94/288157. Yet a further example of an AAV vector employable in this invention is  
20 SSV9AFABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470. Additional AAV gene therapy vectors are described in US 5,354,678, US 5,173,414, US 5,139,941, and US 5,252,479.

The gene therapy vectors of the invention also include herpes vectors. Leading and preferred examples are  
25 herpes simplex virus vectors containing a sequence encoding a thymidine kinase polypeptide such as those disclosed in US 5,288,641 and EP0176170 (Roizman). Additional exemplary herpes simplex virus vectors include HFEM/ICP6-LacZ disclosed in WO95/04139 (Wistar), pHSVlac described in Geller (1988) *Science* 241:1667-1669 and in WO90/09441 & WO92/07945, HSV Us3::pgC-lacZ described in Fink (1992) *Human Gene Therapy* 3:11-19 and HSV 7134, 2 RH 105 and GAL4 described in EP 0453242 (Breakefield), and those  
30 deposited with ATCC as accession numbers ATCC VR-977 and ATCC VR-260.

Also contemplated are alpha virus gene therapy vectors that can be employed in this invention. Preferred alpha virus vectors are Sindbis viruses vectors. Togaviruses, Semliki Forest virus (ATCC VR-67; ATCC VR-1247), Middleberg virus (ATCC VR-370), Ross River virus (ATCC VR-373; ATCC VR-1246), Venezuelan equine encephalitis virus (ATCC VR923; ATCC VR-1250; ATCC VR-1249; ATCC VR-532), and those described in  
35 US patents 5,091,309, 5,217,879, and WO92/10578. More particularly, those alpha virus vectors described in US Serial No. 08/405,627, filed March 15, 1995, WO94/21792, WO92/10578, WO95/07994, US 5,091,309 and US 5,217,879 are employable. Such alpha viruses may be obtained from depositories or collections such as the ATCC in Rockville, Maryland or isolated from known sources using commonly available techniques. Preferably, alphavirus vectors with reduced cytotoxicity are used (see USSN 08/679640).

40 DNA vector systems such as eukaryotic layered expression systems are also useful for expressing the nucleic acids of the invention. See WO95/07994 for a detailed description of eukaryotic layered expression systems.

Preferably, the eukaryotic layered expression systems of the invention are derived from alphavirus vectors and most preferably from Sindbis viral vectors.

Other viral vectors suitable for use in the present invention include those derived from poliovirus, for example ATCC VR-58 and those described in Evans, *Nature* 339 (1989) 385 and Sabin (1973) *J. Biol. Standardization* 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) *J Cell Biochem* L401; pox viruses such as canary pox virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) *Proc Natl Acad Sci* 86:317; Flexner (1989) *Ann NY Acad Sci* 569:86, Flexner (1990) *Vaccine* 8:17; in US 4,603,112 and US 4,769,330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those described in Mulligan (1979) *Nature* 277:108 and Madzak (1992) *J Gen Virol* 73:1533; influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) *Proc Natl Acad Sci* 87:3802-3805; Enami & Palese (1991) *J Virol* 65:2711-2713 and Luytjes (1989) *Cell* 59:110, (see also McMichael (1983) *NEJ Med* 309:13, and Yap (1978) *Nature* 273:238 and *Nature* (1979) 277:108); human immunodeficiency virus as described in EP-0386882 and in Buchschacher (1992) *J. Virol.* 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example ATCC VR-600 and ATCC VR-1240; Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924; Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyzylagach virus, for example ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for example ATCC VR-372 and ATCC VR-1245; Tonate virus, for example ATCC VR-925; Trinit virus, for example ATCC VR-469; Una virus, for example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; O'Nyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70, ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus, for example ATCC VR-740 and those described in Hamre (1966) *Proc Soc Exp Biol Med* 121:190.

Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic acid expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see US Serial No. 08/366,787, filed December 30, 1994 and Curiel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eucaryotic cell delivery vehicles cells, for example see US Serial No.08/240,030, filed May 9, 1994, and US Serial No. 08/404,796, deposition of photopolymerized hydrogel materials, hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and in Woffendin (1994) *Proc Natl Acad Sci* 91:1581-1585.

Particle mediated gene transfer may be employed, for example see US Serial No. 60/023,867. Briefly, the sequence can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, as described in Wu & Wu (1987) *J. Biol. Chem.* 262:4429-4432, insulin as described in Hucked (1990) *Biochem Pharmacol* 40:253-263, galactose as described in Plank (1992) *Bioconjugate Chem* 3:533-539, lactose or transferrin.

Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO90/11092 and US 5,580,859. Uptake efficiency may be improved using biodegradable latex beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the beads. The method may be improved further by treatment of the beads to increase hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm.

Liposomes that can act as gene delivery vehicles are described in US 5,422,120, WO95/13796, WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral delivery, the nucleic acid sequences encoding a polypeptide can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then be incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate DNA comprising the gene under the control of a variety of tissue-specific or ubiquitously-active promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al* (1994) *Proc. Natl. Acad. Sci. USA* 91(24):11581-11585. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in US 5,149,655; use of ionizing radiation for activating transferred gene, as described in US 5,206,152 and WO92/11033

Exemplary liposome and polycationic gene delivery vehicles are those described in US 5,422,120 and 4,762,915; in WO 95/13796; WO94/23697; and WO91/14445; in EP-0524968; and in Stryer, *Biochemistry*, pages 236-240 (1975) W.H. Freeman, San Francisco; Szoka (1980) *Biochem Biophys Acta* 600:1; Bayer (1979) *Biochem Biophys Acta* 550:464; Rivnay (1987) *Meth Enzymol* 149:119; Wang (1987) *Proc Natl Acad Sci* 84:7851; Plant (1989) *Anal Biochem* 176:420.

A polynucleotide composition can comprises therapeutically effective amount of a gene therapy vehicle, as the term is defined above. For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

#### Delivery Methods

Once formulated, the polynucleotide compositions of the invention can be administered (1) directly to the subject; (2) delivered *ex vivo*, to cells derived from the subject; or (3) *in vitro* for recombinant protein expression. The subjects to be treated can be mammals or birds. Also, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (*e.g.* see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in *e.g.* WO93/14778. Examples of cells useful in *ex vivo* applications include, for example, stem cells, particularly hematopoietic, lymph cells, macrophages, dendritic cells, or tumor cells.

Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by the following procedures, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

5 Polynucleotide and polypeptide pharmaceutical compositions

In addition to the pharmaceutically acceptable carriers and salts described above, the following additional agents can be used with polynucleotide and/or polypeptide compositions.

A. Polypeptides

10 One example are polypeptides which include, without limitation: asioloorosomucoid (ASOR); transferrin; asialoglycoproteins; antibodies; antibody fragments; ferritin; interleukins; interferons, granulocyte, macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor and erythropoietin. Viral antigens, such as envelope proteins, can also be used. Also, proteins from other invasive organisms, such as the 17 amino acid peptide from the circumsporozoite protein of plasmodium falciparum known as RII.

15 B. Hormones, Vitamins, etc.

Other groups that can be included are, for example: hormones, steroids, androgens, estrogens, thyroid hormone, or vitamins, folic acid.

C. Polyalkylenes, Polysaccharides, etc.

20 Also, polyalkylene glycol can be included with the desired polynucleotides/polypeptides. In a preferred embodiment, the polyalkylene glycol is polyethylene glycol. In addition, mono-, di-, or polysaccharides can be included. In a preferred embodiment of this aspect, the polysaccharide is dextran or DEAE-dextran. Also, chitosan and poly(lactide-co-glycolide)

D. Lipids, and Liposomes

25 The desired polynucleotide/polypeptide can also be encapsulated in lipids or packaged in liposomes prior to delivery to the subject or to cells derived therefrom.

Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed polynucleotide to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim. Biophys. Acta.* 1097:1-17; Straubinger (1983) *Meth.*  
30 *Enzymol.* 101:512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner (1987) *Proc. Natl. Acad. Sci. USA* 84:7413-7416); mRNA (Malone (1989) *Proc. Natl. Acad. Sci. USA* 86:6077-6081); and purified transcription factors (Debs (1990) *J. Biol. Chem.*  
35 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1-2,3-dioleoyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See,



also, Felgner *supra*). Other commercially available liposomes include transfectace (DDAB/DOPE) and DOTAP/DOPE (Boehringer). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, e.g. Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; WO90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilammellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See e.g. Straubinger (1983) *Meth. Immunol.* 101:512-527; Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; Papahadjopoulos (1975) *Biochim. Biophys. Acta* 394:483; Wilson (1979) *Cell* 17:77; Deamer & Bangham (1976) *Biochim. Biophys. Acta* 443:629; Ostro (1977) *Biochem. Biophys. Res. Commun.* 76:836; Fraley (1979) *Proc. Natl. Acad. Sci. USA* 76:3348; Enoch & Strittmatter (1979) *Proc. Natl. Acad. Sci. USA* 76:145; Fraley (1980) *J. Biol. Chem.* (1980) 255:10431; Szoka & Papahadjopoulos (1978) *Proc. Natl. Acad. Sci. USA* 75:145; and Schaefer-Ridder (1982) *Science* 215:166.

#### E. Lipoproteins

In addition, lipoproteins can be included with the polynucleotide/polypeptide to be delivered. Examples of lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL. Mutants, fragments, or fusions of these proteins can also be used. Also, modifications of naturally occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with the polynucleotide to be delivered, no other targeting ligand is included in the composition.

Naturally occurring lipoproteins comprise a lipid and a protein portion. The protein portion are known as apoproteins. At the present, apoproteins A, B, C, D, and E have been isolated and identified. At least two of these contain several proteins, designated by Roman numerals, AI, AII, AIV; CI, CII, CIII.

A lipoprotein can comprise more than one apoprotein. For example, naturally occurring chylomicrons comprises of A, B, C, & E, over time these lipoproteins lose A and acquire C and E apoproteins. VLDL comprises A, B, C, & E apoproteins, LDL comprises apoprotein B; HDL comprises apoproteins A, C, & E.

The amino acid of these apoproteins are known and are described in, for example, Breslow (1985) *Annu Rev. Biochem* 54:699; Law (1986) *Adv. Exp Med. Biol.* 151:162; Chen (1986) *J Biol Chem* 261:12918; Kane (1980) *Proc Natl Acad Sci USA* 77:2465; and Utermann (1984) *Hum Genet* 65:232.

Lipoproteins contain a variety of lipids including, triglycerides, cholesterol (free and esters), and phospholipids. The composition of the lipids varies in naturally occurring lipoproteins. For example, chylomicrons comprise mainly triglycerides. A more detailed description of the lipid content of naturally occurring lipoproteins can be found, for example, in *Meth. Enzymol.* 128 (1986). The composition of the lipids are chosen to aid in

conformation of the apoprotein for receptor binding activity. The composition of lipids can also be chosen to facilitate hydrophobic interaction and association with the polynucleotide binding molecule.

Naturally occurring lipoproteins can be isolated from serum by ultracentrifugation, for instance. Such methods are described in *Meth. Enzymol.* (*supra*); Pitas (1980) *J. Biochem.* 255:5454-5460 and Mahey (1979) *J Clin. Invest* 64:743-750. Lipoproteins can also be produced by *in vitro* or recombinant methods by expression of the apoprotein genes in a desired host cell. See, for example, Atkinson (1986) *Annu Rev Biophys Chem* 15:403 and Radding (1958) *Biochim Biophys Acta* 30: 443. Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Technologies, Inc., Stoughton, Massachusetts, USA. Further description of lipoproteins can be found in Zuckermann *et al.* PCT/US97/14465.

#### 10 F. Polycationic Agents

Polycationic agents can be included, with or without lipoprotein, in a composition with the desired polynucleotide/polypeptide to be delivered.

Polycationic agents, typically, exhibit a net positive charge at physiological relevant pH and are capable of neutralizing the electrical charge of nucleic acids to facilitate delivery to a desired location. These agents have both *in vitro*, *ex vivo*, and *in vivo* applications. Polycationic agents can be used to deliver nucleic acids to a living subject either intramuscularly, subcutaneously, etc.

The following are examples of useful polypeptides as polycationic agents: polylysine, polyarginine, polyornithine, and protamine. Other examples include histones, protamines, human serum albumin, DNA binding proteins, non-histone chromosomal proteins, coat proteins from DNA viruses, such as (X174, transcriptional factors also contain domains that bind DNA and therefore may be useful as nucleic acid condensing agents. Briefly, transcriptional factors such as C/EBP, c-jun, c-fos, AP-1, AP-2, AP-3, CPF, Prot-1, Sp-1, Oct-1, Oct-2, CREP, and TFIID contain basic domains that bind DNA sequences.

Organic polycationic agents include: spermine, spermidine, and putrescine.

The dimensions and of the physical properties of a polycationic agent can be extrapolated from the list above, to construct other polypeptide polycationic agents or to produce synthetic polycationic agents.

Synthetic polycationic agents which are useful include, for example, DEAE-dextran, polybrene. Lipofectin™, and lipofectAMINE™ are monomers that form polycationic complexes when combined with polynucleotides/polypeptides.

#### Nucleic Acid Hybridisation

"Hybridization" refers to the association of two nucleic acid sequences to one another by hydrogen bonding. Typically, one sequence will be fixed to a solid support and the other will be free in solution. Then, the two sequences will be placed in contact with one another under conditions that favor hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt's reagent or BLOTTO); concentration of the sequences; use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.* [*supra*] vol.2, chapt.9, pp.9.47 to 9.57.

"Stringency" refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated  $T_m$  of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA  
5 immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1 µg for a plasmid or phage digest to  $10^{-9}$  to  
10  $10^{-8}$  g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 µg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of  $10^8$  cpm/µg. For a single-copy mammalian gene a conservative approach would start with  
15 10 µg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than  $10^8$  cpm/µg, resulting in an exposure time of ~24 hours.

Several factors can affect the melting temperature ( $T_m$ ) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length  
20 and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

$$T_m = 81 + 16.6(\log_{10} C_i) + 0.4[\%(G + C)] - 0.6(\% \text{ formamide}) - 600/n - 1.5(\% \text{ mismatch}).$$

where  $C_i$  is the salt concentration (monovalent ions) and  $n$  is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) *Anal. Biochem.* 138: 267-284).

25 In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of the hybridization increases (*ie.* stringency), it becomes less likely for hybridization to occur between strands that are nonhomologous, and as a result, background decreases. If the radiolabeled probe is not completely homologous with the immobilized fragment (as is frequently the case in  
30 gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of the washes affects the intensity of the hybridizing band and the degree of background in a similar manner. The stringency of the washes is also increased with decreasing salt concentrations.

In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with  
35 is 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology, and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background are observed after autoradiography, the filter can be washed at high stringency and

reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

#### Nucleic Acid Probe Assays

5 Methods such as PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes according to the invention can determine the presence of cDNA or mRNA. A probe is said to "hybridize" with a sequence of the invention if it can form a duplex or double stranded complex, which is stable enough to be detected.

10 The nucleic acid probes will hybridize to the Chlamydial nucleotide sequences of the invention (including both sense and antisense strands). Though many different nucleotide sequences will encode the amino acid sequence, the native Chlamydial sequence is preferred because it is the actual sequence present in cells. mRNA represents a coding sequence and so a probe should be complementary to the coding sequence; single-stranded cDNA is complementary to mRNA, and so a cDNA probe should be complementary to the non-coding sequence.

15 The probe sequence need not be identical to the Chlamydial sequence (or its complement) — some variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe can form a duplex with target nucleotides, which can be detected. Also, the nucleic acid probe can include additional nucleotides to stabilize the formed duplex. Additional Chlamydial sequence may also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence may be attached to the 5' end of the probe, with the remainder of the probe sequence being complementary to a Chlamydial sequence. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe  
20 sequence has sufficient complementarity with the a Chlamydial sequence in order to hybridize therewith and thereby form a duplex which can be detected.

The exact length and sequence of the probe will depend on the hybridization conditions, such as temperature, salt condition and the like. For example, for diagnostic applications, depending on the complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more  
25 preferably  $\geq 30$  nucleotides, although it may be shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

Probes may be produced by synthetic procedures, such as the triester method of Matteucci *et al.* [*J. Am. Chem. Soc.* (1981) 103:3185], or according to Urdea *et al.* [*Proc. Natl. Acad. Sci. USA* (1983) 80: 7461], or using commercially available automated oligonucleotide synthesizers.

30 The chemical nature of the probe can be selected according to preference. For certain applications, DNA or RNA are appropriate. For other applications, modifications may be incorporated *e.g.* backbone modifications, such as phosphorothioates or methylphosphonates, can be used to increase *in vivo* half-life, alter RNA affinity, increase nuclease resistance *etc.* [*e.g.* see Agrawal & Iyer (1995) *Curr Opin Biotechnol* 6:12-19; Agrawal (1996) *TIBTECH* 14:376-387]; analogues such as peptide nucleic acids may also be used [*e.g.* see Corey (1997) *TIBTECH* 15:224-229; Buchardt *et al.* (1993) *TIBTECH* 11:384-386].  
35

Alternatively, the polymerase chain reaction (PCR) is another well-known means for detecting small amounts of target nucleic acids. The assay is described in: Mullis *et al.* [*Meth. Enzymol.* (1987) 155: 335-350]; US patents 4,683,195 & 4,683,202. Two 'primers' hybridize with the target nucleic acids and are used to prime the reaction. The primers can comprise sequence that does not hybridize to the sequence of the amplification target (or its

complement) to aid with duplex stability or, for example, to incorporate a convenient restriction site. Typically, such sequence will flank the desired Chlamydial sequence.

A thermostable polymerase creates copies of target nucleic acids from the primers using the original target nucleic acids as a template. After a threshold amount of target nucleic acids are generated by the polymerase, they can be detected by more traditional methods, such as Southern blots. When using the Southern blot method, the labelled probe will hybridize to the Chlamydial sequence (or its complement).

Also, mRNA or cDNA can be detected by traditional blotting techniques described in Sambrook *et al* [*supra*]. mRNA, or cDNA generated from mRNA using a polymerase enzyme, can be purified and separated using gel electrophoresis. The nucleic acids on the gel are then blotted onto a solid support, such as nitrocellulose. The solid support is exposed to a labelled probe and then washed to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected. Typically, the probe is labelled with a radioactive moiety.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1-189 show data pertaining to examples 1-189.

Figure 190 shows a representative 2D gel of proteins in elementary bodies.

Figure 191 shows an alignment of sequences in five (six) proteins of the invention.

## EXAMPLES

The examples indicate *C.pneumoniae* proteins, together with evidence to support the view that the proteins are useful antigens for vaccine production and development or for diagnostic purposes. This evidence takes the form of:

- Computer prediction based on sequence information from CWL029 strain (*e.g.* using the PSORT algorithm available from [www.psорт.nibb.ac.jp](http://www.psорт.nibb.ac.jp)).
- Data on recombinant expression and purification of the proteins cloned from IOL207 strain.
- Western blots to demonstrate immunoreactivity in serum (typically a blot of an EB extract of *C.pneumoniae* strain FB/96 stained with mouse antiserum against the recombinant protein).
- FACS analysis of *C.pneumoniae* bacteria or purified EBs to confirm accessibility of the antigen to the immune system (see also table III).
- An indication if the protein was identified by MALDI-TOF from a 2D gel electrophoresis map of proteins from purified elementary bodies from strain FB/96. This confirms that the protein is expressed *in vivo* (see also table V).

Various tests can be used to assess the *in vivo* immunogenicity of the proteins identified in the examples. For example, the proteins can be expressed recombinantly and used to screen patient sera by immunoblot. A positive reaction between the protein and patient serum indicates that the patient has previously mounted an immune response to the protein in question *ie.* the protein is an immunogen. This method can also be used to identify immunodominant proteins.

## CLONING OF CPN ORFs FOR EXPRESSION IN *E. COLI*

10

- a) proteins having an hexa-histidine tag at the C-terminus (cpn-His)
- b) proteins having a GST fusion partner at the N-terminus (Gst-cpn)
- c) proteins having both hexa-histidine tag at the C-terminus and GST at the N-terminus (GST/His fusion; NH<sub>2</sub>-GST-cpn-(His)<sub>6</sub>-COOH)

20 The modified versions of pGEX-KG vector were made with the aim of allowing the cloning of single amplification products in all three vectors after only one double restriction enzyme digestion and to minimise the presence of extraneous amino acids in the final recombinant proteins.

Two couples of complementary oligodeoxyribonucleotides were synthesised using the DNA synthesiser ABI394 (Perkin Elmer) and the reagents from Cruachem (Glasgow, Scotland). Equimolar amounts of the oligo pairs (50 ng each oligo) were annealed in T4 DNA ligase buffer (New England Biolabs) for 10 min in a final volume of 50µl and then were left to cool slowly at room temperature. With the described procedure the following DNA linkers were obtained:

30 NdeI NheI XmaI EcoRI NcoI SalI XhoI SacI NotI  
GATCCCATATGGCTAGCCCGGGAATTCTGTCATGGAGTGAGTCGACTGACTCGAGTGATCGAGCTCCTGAGCGGCCGCATGAA  
GGTATACCGATCGGGCCCCCTTAAGCAGGTACCTCACTCAGCTGACTGAGCTCACTAGCTCGAGGACTCGCCGGCGTACTTTCGA

35 HindIII NotI XhoI --Hexa-Histidine--  
TCGACAAGCTTGC GGCCGCACTCGAGCATCACCATCACCATCACTGAT  
GTTCCAACGCCGGCGTGAGCACGTAGAGGTAGTGGTAGTGACTATCGA

The plasmid pGEX-KG was digested with BamHI and HindIII and 100 ng were ligated overnight at 16 °C to the linker gexNN with a molar ratio of 3:1 linker/plasmid using 200 units of T4 DNA ligase

(New England Biolabs). After transformation of the ligation product in *E. coli* DH5, a clone containing the pGEX-NN plasmid, having the correct linker, was selected by means of restriction enzyme analysis and DNA sequencing.

The new plasmid pGEX-NN was digested with Sall and HindIII and ligated to the linker gexNNH.

- 5 After transformation of the ligation product in *E. coli* DH5, a clone containing the pGEX-NNH plasmid, having the correct linker, was selected by means of restriction enzyme analysis and DNA sequencing.

### (B) Chromosomal DNA preparation

- 10 The chromosomal DNA of elementary bodies (EB) of *C. pneumoniae* strain 10L-207 was prepared by adding 1.5 ml of lysis buffer (10 mM Tris-HCl, 150 mM NaCl, 2 mM EDTA, 0.6 % SDS, 100 µg/ml Proteinase K, pH 8) to 450 µl EB suspension (400.000/µl) and incubating overnight at 37 °C. After sequential extraction with phenol, phenol-chloroform, and chloroform, the DNA was precipitated with 0.3 M sodium acetate, pH 5.2 and 2 volumes of absolute ethanol. The DNA pellet was washed with 70 % ethanol. After solubilization with distilled water and treatment with 20 µg/ml RNase A
- 15 for 1 hour at RT, the DNA was extracted again with phenol-chloroform, alcohol precipitated and suspended with 300 µl 1 mM Tris-HCl pH 8.5. The DNA concentration was evaluated by measuring OD<sub>260</sub> of the sample.

### (C) Oligonucleotide design

- 20 Synthetic oligonucleotide primers were designed on the basis of the coding sequence of each ORF using the sequence of *C. pneumoniae* strain CWL029. Any predicted signal peptide were omitted, by deducing the 5' end amplification primer sequence immediately downstream from the predicted leader sequence. For most ORFs, the 5' tail of the primers (table I) included only one restriction enzyme recognition site (NdeI, or NheI, or SpeI depending on the gene's own restriction pattern); the 3' primer tails (table I) included a XhoI or a NotI or a HindIII restriction site.

5' tails		3' tails	
NdeI	5' GTGCGTCATATG 3'	XhoI	5' GCGTCTCGAG 3'
NheI	5' GTGCGTGCTAGC 3'	NotI	5' ACTCGCTAGCGGCCGC 3'
SpeI	5' GTGCGTACTAGT 3'	HindIII	5' GCGTAAGCTT 3'

25 **Table I.** Oligonucleotide tails of the primers used to amplify Cpn genes.

- As well as containing the restriction enzyme recognition sequences, the primers included nucleotides which hybridized to the sequence to be amplified. The number of hybridizing nucleotides depended on the melting temperature of the primers which was determined as described [(Breslauer *et al.* (1986) *PNAS USA* 83:3746-50)]. The average melting temperature of the selected oligos was 50-55°C
- 30 for the hybridizing region alone and 65-75°C for the whole oligos. Table II shows the forward and reverse primers used for each amplification.

**(D) Amplification**

The standard PCR protocol was as follow: 50 ng genomic DNA were used as template in the presence of 0,2  $\mu$ M each primer, 200  $\mu$ M each dNTP, 1,5 mM  $MgCl_2$ , 1x PCR buffer minus Mg (Gibco-BRL), and 2 units of Taq DNA polymerase (Platinum Taq, Gibco-BRL) in a final volume of 100  $\mu$ l. Each sample underwent a double-step amplification: the first 5 cycles were performed using as the hybridizing temperature the one of the oligos excluding the restriction enzyme tail, followed by 25 cycles performed according to the hybridization temperature of the whole lenght primers. The standard cycles were as follow:

denaturation : 94 °C, 2 min

denaturation: 94 °C, 30 seconds	}	5 cycles
hybridization: 51 °C, 50 seconds		
elongation: 72 °C, 1 min or 2 min and 40 sec		

denaturation: 94 °C, 30 seconds	}	25 cycles
hybridization: 70 °C, 50 seconds		
elongation: 72 °C, 1 min or 2 min and 40 sec		

72 °C, 7 min

4 °C

The elongation time was 1 min for ORFs shorter than 2000 bp, and 2 min and 40 seconds for ORFs longer than 2000 bp. The amplifications were performed using a Gene Amp PCR system 9600 (Perkin Elmer).

To check the amplification results, 4  $\mu$ l of each PCR product was loaded onto 1-1.5 agarose gel and the size of amplified fragments compared with DNA molecular weight standards (DNA markers III or IX, Roche). The PCR products were loaded on agarose gel and after electrophoresis the right size bands were excised from the gel. The DNA was purified from the agarose using the Gel Extraction Kit (Qiagen) following the instruction of the manufacturer. The final elution volume of the DNA was 50  $\mu$ l TE (10 mM Tris-HCl, 1 mM EDTA, pH 8). One  $\mu$ l of each purified DNA was loaded onto agarose gel to evaluate the yield.

**(E) Digestion of PCR fragments**

One-two  $\mu$ g of purified PCR product were double digested overnight at 37 °C with the appropriate restriction enzymes (60 units of each enzyme) using the appropriate restriction buffer in 100  $\mu$ l final volume. The restriction enzymes and the digestion buffers were from New England Biolabs. After



purification of the digested DNA (PCR purification Kit, Qiagen) and elution with 30 µl TE, 1 µl was subjected to agarose gel electrophoresis to evaluate the yield in comparison to titrated molecular weight standards (DNA markers III or IX, Roche).

**(F) Digestion of the cloning vectors (pET21b+, pGEX-NN, and pGEX-NNH)**

- 5 10 µg of plasmid was double digested with 100 units of each restriction enzyme in 400 µl reaction volume in the presence of appropriate buffer by overnight incubation at 37 °C. After electrophoresis on a 1% agarose gel, the band corresponding to the digested vector was purified from the gel using the Qiagen Qiaex II Gel Extraction Kit and the DNA was eluted with 50 µl TE. The DNA concentration was evaluated by measuring OD<sub>260</sub> of the sample.

10 **(G) Cloning**

75ng of the appropriately digested and purified vectors and the digested and purified fragments corresponding to each ORF, were ligated in final volumes of 10-20 µl with a molar ratio of 1:1 fragment/vector, using 400 units T4 DNA ligase (New England Biolabs) in the presence of the buffer supplied by the manufacturer. The reactions were incubated overnight at 16 °C.

- 15 Transformation in *E. coli* DH5 competent cells was performed as follow: the ligation reaction was mixed with 200 µl of competent DH5 cells and incubated on ice for 30 min and then at 42 °C for 90 seconds. After cooling on ice, 0.8 ml LB was added and the cells were incubated for 45 min at 37 °C under shaking. 100 and 900 µl of cell suspensions were plated on separate plates of agar LB 100 µg/ml Ampicillin and the plates were incubated overnight at 37 °C. The screening of the  
20 transformants was done by growing randomly chosen clones in 6 ml LB 100 µg/ml Ampicillin, by extracting the DNA using the Qiagen Qiaprep Spin Miniprep Kit following the manufacturer instructions, and by digesting 2 µl of plasmid miniprep with the restriction enzymes specific for the restriction cloning sites. After agarose gel electrophoresis of the digested plasmid mini-preparations, positive clones were chosen on the basis of the correct size of the restriction fragments,  
25 as evaluated by comparison with appropriate molecular weight markers (DNA markers III or IX, Roche).

**(H) Expression**

- 1 µl of each right plasmid mini-preparation was transformed in 200 µl of competent *E. coli* strain suitable for expression of the recombinant protein. All pET21b+ recombinant plasmids were  
30 transformed in BL21 DE3 (Novagen) *E. coli* cells, whilst all pGEX-NN and all pGEX-NNH recombinant plasmids were transformed in BL21 cells (Novagen). After plating transformation mixtures on LB/Amp agar plates and incubation overnight at 37 °C, single colonies were inoculated in 3 ml LB 100 µg/ml Ampicillin and grown at 37 °C overnight. 70 µl of the overnight culture was inoculated in 2 ml LB/Amp and grown at 37 °C until OD<sub>600</sub> of the pET clones reached the 0,4-0,8  
35 value or until OD<sub>600</sub> of the pGEX clones reached the 0,8-1 value. Protein expression was then

induced by adding IPTG (Isopropil  $\beta$ -D thio-galacto-piranoside) to the mini-cultures. pET clones were induced using 1 mM IPTG, whilst pGEX clones were induced using 0.2 mM IPTG. After 3 hours incubation at 37 °C the final OD<sub>600</sub> was checked and the cultures were cooled on ice. After centrifugation of 0.5 ml culture, the cell pellet was suspended in 50  $\mu$ l of protein Loading Sample Buffer (60 mM TRIS-HCl pH 6.8, 5% w/v SDS, 10% v/v glycerin, 0.1% w/v Bromophenol Blue, 100 mM DTT) and incubated at 100 °C for 5 min. A volume of boiled sample corresponding to 0.1 OD<sub>600</sub> culture was analysed by SDS-PAGE and Coomassie Blue staining to verify the presence of induced protein band.

## PURIFICATION OF THE RECOMBINANT PROTEINS

Single colonies were inoculated in 25 ml LB 100  $\mu$ g/ml Ampicillin and grown at 37 °C overnight. The overnight culture was inoculated in 500 ml LB/Amp and grown under shaking at 25 °C until OD<sub>600</sub> 0,4-0,8 value for the pET clones, or until OD<sub>600</sub> 0,8-1 value for the pGEX clones. Protein expression was then induced by adding IPTG to the cultures. pET clones were induced using 1 mM IPTG, whilst pGEX clones were induced using 0.2 mM IPTG. After 4 hours incubation at 25 °C the final OD<sub>600</sub> was checked and the cultures were cooled on ice. After centrifugation at 6000 rpm (JA10 rotor, Beckman), the cell pellet was processed for purification or frozen at -20 °C.

### (I) Procedure for the purification of soluble His-tagged proteins from *E.coli*

1. Transfer the pellets from -20°C to ice bath and reconstitute with 10 ml 50 mM NaHPO<sub>4</sub> buffer, 300 mM NaCl, pH 8,0, pass in 40-50 ml centrifugation tubes and break the cells as per the following outline:
2. Break the pellets in the French Press performing three passages with in-line washing.
3. Centrifuge at about 30-40000 x g per 15-20 min. If possible use rotor JA 25.50 (21000 rpm, 15 min.) or JA-20 (18000 rpm, 15 min.)
4. Equilibrate the Poly-Prep columns with 1 ml Fast Flow Chelating Sepharose resin with 50 mM phosphate buffer, 300 mM NaCl, pH 8,0.
5. Store the centrifugation pellet at -20°C, and load the supernatant in the columns.
6. Collect the flow through.
7. Wash the columns with 10 ml (2 ml + 2 ml + 4 ml) 50 mM phosphate buffer, 300 mM NaCl, pH 8,0.
8. Wash again with 10 ml 20 mM imidazole buffer, 50 mM phosphate, 300 mM NaCl, pH 8,0.
9. Elute the proteins bound to the columns with 4,5 ml (1,5 ml + 1,5 ml + 1,5 ml) 250 mM imidazole buffer, 50 mM phosphate, 300 mM NaCl, pH 8,0 and collect the 3 corresponding fractions of ~1,5 ml each. Add to each tube 15  $\mu$ l DTT 200 mM (final concentration 2 mM)

10. Measure the protein concentration of the first two fractions with the Bradford method, collect a 10 µg aliquot of proteins from each sample and analyse by SDS-PAGE. (N.B.: should the sample be too diluted, load 21 µl + 7 µl loading buffer).

11. Store the collected fractions at +4°C while waiting for the results of the SDS-PAGE analysis.

5 12. For immunisation prepare 4-5 aliquots of 100 µg each in 0,5 ml in 40% glycerol. The dilution buffer is the above elution buffer, plus 2 mM DTT. Store the aliquots at -20°C until immunisation.

#### (J) Purification of His-tagged proteins from Inclusion bodies

Purifications were carried out essentially according the following protocol:

10 1. Bacteria are collected from 500 ml cultures by centrifugation. If required store bacterial pellets at -20°C. For extraction, resuspend each bacterial pellet in 10 ml 50 mM TRIS-HCl buffer, pH 8,5 on an ice bath.

2. Disrupt the resuspended bacteria with a French Press, performing two passages.

15 3. Centrifuge at 35000 x g for 15 min and collect the pellets. Use a Beckman rotor JA 25.50 (21000 rpm, 15 min.) or JA-20 (18000 rpm, 15 min.).

4. Dissolve the centrifugation pellets with 50 mM TRIS-HCl, 1 mM TCEP {Tris(2-carboxyethyl)-phosphine hydrochloride, Pierce} , 6M guanidium chloride, pH 8,5. Stir for ~ 10 min. with a magnetic bar.

5. Centrifuge as described above, and collect the supernatant..

20 6. Prepare an adequate number of Poly-Prep (Bio-Rad) columns containing 1 ml of Fast Flow Chelating Sepharose (Pharmacia) saturated with Nichel according to manufacturer recommendations.. Wash the columns twice with 5 ml of H<sub>2</sub>O and equilibrate with 50 mM TRIS-HCl, 1 mM TCEP, 6M guanidinium chloride, pH 8,5.

25 7. Load the supernatants from step 5 onto the columns, and wash with 5 ml of 50 mM TRIS-Hcl buffer, 1 mM TCEP, 6M urea, pH 8,5

8. Wash the columns with 10 ml of 20 mM imidazole, 50 mM TRIS-HCl , 6M urea, 1 mM TCEP, pH 8,5. Collect and set aside the first 5 ml for possible further controls.

30 9. Elute the proteins bound to the columns with 4,5 ml of a buffer containing 250 mM imidazole, 50 mM TRIS-HCl, 6M urea, 1 mM TCEP, pH 8,5. Add the elution buffer in three 1,5 ml aliquots, and collect the corresponding 3 fractions. Add to each fraction 15 µl DTT (final concentration 2 mM) .

10. Measure eluted protein concentration with the Bradford method, and analyze aliquots of ca 10 µg of protein by SDS-PAGE.

35 11. Store proteins at -20°C in 40% (v/v) glycerol, 50 mM TRIS-HCl, 2M urea, 0.5 M arginine, 2 mM DTT, 0.3 mM TCEP, 83.3 mM imidazole, pH 8,5

**(K) Procedure for the purification of GST-fusion proteins from *E.coli***

1. Transfer the bacterial pellets from  $-20^{\circ}\text{C}$  to an ice bath and resuspend with 7,5 ml PBS, pH 7,4 to which a mixture of protease inhibitors (CØMPLETE™ - Boehringer Mannheim, 1 tablet every 25 ml of buffer) has been added. Transfer to 40-50 ml centrifugation tubes and sonicate according to the following procedure:
  - a) Position the probe at about 0,5 cm from the bottom of the tube
  - b) Block the tube with the clamp
  - c) Dip the tube in an ice bath
  - d) Set the sonicator as follows: Timer  $\rightarrow$  Hold, Duty Cycle  $\rightarrow$  55, Out. Control  $\rightarrow$  6.
  - e) perform 5 cycles of 10 impulses at a time lapse of 1 minute (i.e. one cycle = 10 impulses + ~45" hold; b. 10 impulses + ~45" hold; c. 10 impulses + ~45" hold; d. 10 impulses + ~45" hold; e. 10 impulses + ~45" hold)
2. Centrifuge at about 30-40000 x g for 15-20 min. E.g.: use rotor Beckman JA 25.50 at 21000 rpm, for 15 min.
3. Store the centrifugation pellets at  $-20^{\circ}\text{C}$ , and load the supernatants on the chromatography columns, as follows
4. Equilibrate the Poly-Prep (Bio-Rad) columns with 0,5 ml ( $\cong$  1 ml suspension) of Glutathione-Sepharose 4B resin, wash with 2 ml (1 + 1)  $\text{H}_2\text{O}$ , and then with 10 ml (2 + 4 + 4) PBS, pH 7,4.
5. Load the supernatants on the columns and discard the flow through.
6. Wash the columns with 10 ml (2 + 4 + 4) PBS, pH 7,4.
7. Elute the proteins bound to the columns with 4,5 ml of 50 mM TRIS buffer, 10 mM reduced glutathione, pH 8,0, adding 1,5 ml + 1,5 ml + 1,5 ml and collecting the respective 3 fractions of ~1,5 ml each.
8. Measure the protein concentration of the first two fractions with the Bradford method, analyse a 10  $\mu\text{g}$  aliquot of proteins from each sample by SDS-PAGE. (N.B.: if the sample is too diluted load 21  $\mu\text{l}$  (+ 7  $\mu\text{l}$  loading buffer).
9. Store the collected fractions at  $+4^{\circ}\text{C}$  while waiting for the results of the SDS-PAGE analysis.
10. For each protein destined to the immunisation prepare 4-5 aliquots of 100  $\mu\text{g}$  each in 0,5 ml of 40% glycerol. The dilution buffer is 50 mM TRIS.HCl, 2 mM DTT, pH 8,0. Store the aliquots at  $-20^{\circ}\text{C}$  until immunisation..

**SEROLOGY****(L) Protocol of immunization**

1. Groups of four CD1 female mice aged between 6 and 7 weeks were immunized with 20  $\mu\text{g}$  of recombinant protein resuspended in 100  $\mu\text{l}$ .

2. Four mice for each group received 3 doses with a 14 days interval schedule.
3. Immunization was performed through intra-peritoneal injection of the protein with an equal volume of Complete Freund's Adjuvant (CFA) for the first dose and Incomplete Freund's Adjuvant (IFA) for the following two doses.
- 5 4. Sera were collected before each immunization. Mice were sacrificed 14 days after the third immunization and the collected sera were pooled and stored at  $-20^{\circ}\text{C}$ .

**(M) Western blot analysis of Cpn elementary body proteins with mouse sera**

Aliquots of elementary bodies containing approximately 4  $\mu\text{g}$  of proteins, mixed with SDS loading buffer (1x: 60 mM TRIS-HCl pH 6.8, 5% w/v SDS, 10% v/v glycerin, 0.1% Bromophenol Blue, 100 mM DTT) and boiled 5 minutes at  $95^{\circ}\text{C}$ , were loaded on a 12% SDS-PAGE gel. The gel was run using a SDS-PAGE running buffer containing 250 mM TRIS, 2.5 mM Glycine and 0.1 %SDS. The gel was electroblotted onto nitrocellulose membrane at 200 mA for 30 minutes. The membrane was blocked for 30 minutes with PBS, 3% skimmed milk powder and incubated O/N at  $4^{\circ}\text{C}$  with the appropriate dilution (1/100) of the sera. After washing twice with PBS + 0.1% Tween (Sigma) the membrane was incubated for 2 hours with peroxidase-conjugated secondary anti-mouse antibody (Sigma) diluted 1:3000. The nitrocellulose was washed twice for 10 minutes with PBS + 0.1% Tween-20 and once with PBS and thereafter developed by Opti-4CN Substrate Kit (Biorad).

Lanes shown in Western blots are: (P) = pre-immune control serum; (I) = immune serum.

**(N) FACS analysis of *Chlamydia pneumoniae* elementary bodies with mouse sera**

- 20 1.  $2 \times 10^5$  Elementary Bodies (EB)/well were washed with 200  $\mu\text{l}$  of PBS-0.1%BSA in a 96 wells U bottom plate and centrifuged for 10 min. at 1200rpm, at  $4^{\circ}\text{C}$ .
2. The supernatant was discarded and the E.B. resuspended in 10  $\mu\text{l}$  of PBS-0.1%BSA.
3. 10 $\mu\text{l}$  mouse sera diluted in PBS-0.1%BSA were added to the E.B. suspension to a final dilution of 1:400, and incubated on ice for 30 min.
- 25 4. EB were washed by adding 180 $\mu\text{l}$  PBS-0.1%BSA and centrifuged for 10min. at 1200rpm,  $4^{\circ}\text{C}$ .
5. The supernatant was discarded and the E.B. resuspended in 10 l of PBS-0.1%BSA.
6. 10 $\mu\text{l}$  of a goat anti-mouse IgG, F(ab')<sub>2</sub> fragment specific-R-Phycoerythrin-conjugated (Jackson Immunoresearch Laboratories Inc., cat.N°115-116-072) was added to the EB suspension to a final dilution of 1:100, and incubated on ice for 30 min. in the dark.
- 30 7. EB were washed by adding 180 $\mu\text{l}$  PBS-0.1%BSA and centrifuged for 10min. at 1200rpm,  $4^{\circ}\text{C}$ .
8. The supernatant was discarded and the E.B. resuspended in 150  $\mu\text{l}$  of PBS-0.1%BSA.
9. E.B. suspension was passed through a cytometric chamber of a FACS Calibur (Becton Dickinson, Mountain View, CA USA) and 10.000 events were acquired.

10. Data were analysed using Cell Quest Software (Becton Dickinson, Mountain View, CA USA) by drawing a morphological dot plot (using forward and side scatter parameters) on E.B. signals. An histogram plot was then created on FL2 intensity of fluorescence log scale recalling the morphological region of EB.

- 5 NB: the results of FACS depend not only on the extent of accessibility of the native antigens but also on the quality of the antibodies elicited by the recombinant antigens, which may have structures with a variable degree of correct folding as compared with the native protein structures. Therefore, even if a FACS assay appears negative this does not necessarily mean that the protein is not abundant or accessible on the surface. PorB antigen, for instance, gave negative results in FACS but is a surface-exposed neutralising antigen [Kubo & Stephens (2000) *Mol. Microbiol.* 38:772-780].

#### (O) Mass Spectrometry analysis of two-dimensional electrophoretic protein maps

Gradient purified EBs from strain FB/96 were solubilized at a final concentration of 5.5mg/ml with immobiline rehydration buffer (7M urea, 2M thiourea, 2% (w/v) CHAPS, 2% (w/v) ASB 14 [Chevallet *et al.* (1998) *Electrophor.* 19:1901-9], 2% (v/v) C.A 3-10NL (Amersham Pharmacia Biotech), 2 mM tributyl phosphine, 65 mM DTT). Samples (250µg protein) were adsorbed overnight on Immobiline DryStrips (7 cm, pH 3-10 non linear). Electrophocusing was performed in a IPGphor Isoelectric Focusing Unit (Amersham Pharmacia Biotech). Before PAGE separation, the focused strips were incubated in 4M urea, 2M thiourea, 30% (v/v) glycerol, 2% (w/v) SDS, 5mM tributyl phosphine 2.5%(w/v) acrylamide, 50mM Tris-HCl pH 8.8, as described [Herbert *et al.* (1998) *Electrophor.* 19:845-51]. SDS-PAGE was performed on linear 9-16% acrylamide gradients. Gels were stained with colloidal Coomassie (Novex, San Diego) [Doherty *et al.* (1998) *Electrophor.* 19:355-63]. Stained gels were scanned with a Personal Densitometer SI (Molecular Dynamics) at 8 bits and 50µm per pixel. Map images were annotated with the software Image Master 2D Elite, version 3.10 (Amersham Pharmacia Biotech). Protein spots were excised from the gel, using an Ettan Spot picker (Amersham Pharmacia Biotech), and dried in a vacuum centrifuge. In-gel digestion of samples for mass spectrometry and extraction of peptides were performed as described by Wilm *et al.* [Nature (1996) 379:466-9]. Samples were desalted with a ZIP TIP (Millipore), eluted with a saturated solution of alpha-cyano-4-hydroxycinnamic acid in 50% acetonitrile, 0.1% TFA and directly loaded onto a SCOUT 381 multiprobe plate (Bruker). Spectra were acquired on a Bruker Biflex II MALDI-TOF. Spectra were calibrated using a combination of known standard peptides, located in spots adjacent to the samples. Resulting values for monoisotopic peaks were used for database searches using the computer program Mascot (www.matrixscience.com). All searches were performed using an error of 200-500ppm as constraint. A representative gel is shown in Figure 190.

#### Example 1

35 The following *C.pneumoniae* protein (PID 4376552) was expressed <SEQ ID 1; cp6552>:

1 MKKKLSLLVG LIFVLSSCHK EDAQNKIRIV ASPTPHAELL ESLQEEAKDL

51 GIKLKILPVD DYRIPNRIIL DKQVDANYFQ HQAFLDDECE RYDCKGELVV  
 101 IAKVHLEPQA IYSKKHSSLE RLKSQKKLTI AIPVDRTNAQ RALHLLLEECG  
 151 LIVCKGPANL NMTAKDVCGK ENRSINILEV SAPLLVGSLP DVDAAVIPGN  
 201 FAIAANLSPK KDSLCLLEDLS VSKYTNLVVI RSEDVGSPPM IKLQKLFQSP  
 251 SVQHFFDTKY HGNILMTQD NG\*

A predicted signal peptide is highlighted.

The cp6552 nucleotide sequence <SEQ ID 2> is:

1 ATGAAAAAAA AATTATCATT ACTTGTAGGT TTAATTTTGT TTTTGAGTTC  
 51 TTGCCATAAG GAAGATGCTC AGAATAAAAT ACGTATTGTA GCCAGTCCGA  
 101 CACCTCATGC GGAATTATTG GAGAGTTTAC AGGAAGAGGC TAAAGATCTT  
 151 GGAATCAAGC TGAAAATACT TCCAGTAGAT GATTATCGTA TTCCTAATCG  
 201 TTTGCTTTTG GATAACAAG TAGATGCAAA TTACTTTCAA CATCAAGCTT  
 251 TTCTTGATGA CGAATGCGAG CGTTATGATT GTAAGGGTGA ATTAGTTGTT  
 301 ATCGCTAAAG TTCATTTGGA ACCTCAAGCA ATTTATTCTA AGAAACATTC  
 15 351 TTCTTTAGAG CGCTTAAAAA GCCAGAAGAA ACTGACTATA GCGATTCTCG  
 401 TGGATCGTAC GAATGCTCAG CGTGCTCTAC ACTTGTTAGA AGAGTGCCTG  
 451 CTCATTGTTT GCAAAGGGCC TGCTAATTTA AATATGACAG CTAAAGATGT  
 501 CTGTGGGAAA GAAAATAGAA GTATCAACAT ATTAGAGGTG TCAGCTCCTC  
 551 TTCTTGTCGG ATCTCTTCCT GACGTTGATG CTGCTGTCAT TCCTGGAAAT  
 20 601 TTTGCTATAG CAGCAAACCT TTCTCCAAAG AAAGATAGTC TTTGTTTAGA  
 651 GGATCTTTTC GTATCTAAGT ATACAAACCT TGTGTTCATT CGTTCTGAAG  
 701 ACGTAGGTTT TCCTAAAATG ATAAAATTAC AGAAGCTGTT TCAATCTCCT  
 751 TCTGTACAAC ATTTTTTGA TACAAAATAT CATGGGAATA TTTTGACAAT  
 801 GACTCAAGAC AATGTTAG

25 The PSORT algorithm predicts an inner membrane location (0.127).

The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 1A, and also as a GST-fusion. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 1B) and for FACS analysis (Figure 1C).

The cp6552 protein was also identified in the 2D-PAGE experiment (Cpn0278).

30 These experiments show that cp6552 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## Example 2

The following *C.pneumoniae* protein (PID 4376736) was expressed <SEQ ID 3; cp6736>:

1 MKTSIRKFLI STTLAPCFAS TAFTVEVIMP SENFDGSSGK IFPYTTLSDP  
 35 51 RGTLCIFSGD LYIANLDNAI SRTSSSCFSN RAGALQILGK GGVFSFLNIR  
 101 SSADGAAISS VITQNPCLP LSFSGFSQMI FDNCESTSD TSASNVIPHA  
 151 SAIYATTPML FTNNDLSILFQ YNRSAGFGAA IRGTSITIEN TKKSLLFNNG  
 201 GSISNGGALT GSAAINLINN SAPVIFSTNA TGIYGGAIYL TGGSMILTSGN  
 251 LSGVLFVNNS SRSGGAIYAN GNVTFSSNSD LTFQNTASP QNSLPAPTTP  
 40 301 PTPPAVTPLL GYGGAIFCTP PATPPPTGVS LTISGENSVT FLENIASEQG  
 351 GALYGKKISI DSNKSTIFLG NTAGKGAIA IPESGELSLS ANQGDILFNK  
 401 NLSITSGTPT RNSIHFGKDA KFATLGATQG YTLYFYDPIT SDDLAAASAA  
 451 ATVVVNPKAS ADGAYSGTIV FSGETLTATE AATPANATST LNQKLELEGG  
 501 TLALRNGATL NVHNFQDEK SVVIMDAGTT LATTNGANNT DGAITLNLKLV  
 45 551 INLDSLDGK AAVVNVQSTN GALTISGTLG LVKNSQDCCD NHGMFNKDLQ  
 601 QVPILELKAT SNTVTTFDFS LGTNGYQOSP YGYQGTWEFT IDTTHTVTG  
 651 NWKKTGYLPH PERLAPLIPN SLWANVIDLR AVSQASAADG EDVPGKQLSI  
 701 TGITNFFHAN HTGDARSYRH MGGGYLINTY TRITPDAALS LGFGQLFTKS  
 751 KDYLVGHHGS NVYFATVYSN ITKSLFGSSR FFSGGTSRVT YSRSEKVKVT  
 50 801 SYTKLPKGRC SWSNNCWLGE LEGNLPITLS SRIILNLKQII PFVKADEVAYA  
 851 THGGIQENTP EGRIFGHGHL LNVAVPVGVR FGKNSHNRPD FYTIIVAYAP  
 901 DVYRHNPD CD TTLPIGATW TSGNNLTRS TLLVQASSHT SVNDVLEIFG  
 951 HCCGDIRRTS RQYTLDIGSK LRF\*

A predicted signal peptide is highlighted.

The cp6736 nucleotide sequence <SEQ ID 4> is:

```

1  ATGAAAACGT CTATTCGTAA GTTCTTAATT TCTACCACAC TGGCGCCATG
51  TTTTGCTTCA ACAGCGTTTA CTGTAGAAGT TATCATGCCT TCCGAGAACT
101  TTGATGGATC GAGTGGGAAG ATTTTTCCTT ACACAACACT TTCTGATCCT
5   151  AGAGGGACAC TCTGTATTTT TTCAGGGGAT CTCTACATTG CGAATCTTGA
201  TAATGCCATA TCCAGAACCT CTTCCAGTTG CTTTAGCAAT AGGGCGGGAG
251  CACTACAAAT CTTAGGAAAA GGTGGGGTTT TCTCCTTCTT AAATATCCGT
301  TCTTCAGCTG ACGGAGCCGC GATTAGTAGT GTAATCACCC AAAATCCTGA
351  ACTATGTCCC TTGAGTTTTT CAGGATTTAG TCAGATGATC TTCGATAACT
10  401  GTGAATCTTT GACTTCAGAT ACCTCAGCGA GTAATGTCAT ACCTCACGCA
451  TCGGCGAATT ACGCTACAAC GCCCATGCTC TTTACAAACA ATGACTCCAT
501  ACTATTCCAA TACAACCGTT CTGCAGGATT TGGAGCTGCC ATTCGAGGCA
551  CAAGCATCAC AATAGAAAAT ACGAAAAAGA GCCTTCTCTT TAATGGTAAT
601  GGATCCATCT CTAATGGAGG GGCCCTCACG GGATCTGCAG CGATCAACCT
15  651  CATCAACAAT AGCGCTCCTG TGATTTTCTC AACGAATGCT ACAGGGATCT
701  ATGGTGGGGC TATTTACCTT ACCGGAGGAT CTATGCTCAC CTCTGGGAAC
751  CTCTCAGGAG TCTTGTTTCG TAATAATAGC TCGCGCTCAG GAGGCGCTAT
801  CTGCTAATAC GGAAATGTCA CATTTTCCTA TAACAGCGCA CTGACTTTCC
851  AAAACAATAC AGCATCTCCA CAAAACCTCT TACCTGCACC TACACCTCCA
20  901  CCTACACCAC CAGCAGTCAC TCCTTTGTTA GGATATGGAG GCGCCATCTT
951  CTGTACTCCT CCAGCTACCC CCCCACCAAC AGGTGTTAGC CTGACTATAT
1001 CTGGAGAAAA CAGCGTTACA TTCCTAGAAA ACATTGCCTC CGAACAAGGA
1051 GGAGCCCTCT ATGGCAAAAA GATCTCTATA GATTCTAATA AATCTACAAT
1101 ATTTCTTGGA AATACAGCTG GAAAAGGAGG CGCTATTGCT ATTCCCGAAT
25  1151 CTGGGGAGCT CTCTCTATCC GCAAATCAAG GTGATATCCT CTTTAACAAG
1201 AACCTCAGCA TCACTAGTGG GACACCTACT CGCAATAGTA TTCACTTCGG
1251 AAAAGATGCC AAGTTTGCCA CTCTAGGAGC TACGCAAGGC TATACCCTAT
1301 ACTTCTATGA TCCGATTACA TCTGATGATT TATCTGCTGC ATCCGCAGCC
1351 GCTACTGTGG TCGTCAATCC CAAAGCCAGT GCAGATGGTG CGTATTCAGG
30  1401 GACTATTGTC TTTTCAGGAG AAACCCTCAC TGCTACCGAA GCAGCAACCC
1451 CTGCAAAATG TACATCTACA TTAAACCAA AGCTAGAACT TGAAGGCGGT
1501 ACTCTCGCTT TAAGAAACGG TGCTACCTTA AATGTTTATA ACTTCACGCA
1551 AGATGAAAAG TCCGTCGTCA TCATGGATGC AGGGACCACA TTAGCAACTA
35  1601 CAAATGGAGC TAATAATACT GACGGTGCTA TCACCTTAAA CAAGCTTGTA
1651 ATCAATCTGG ATTCTTTTGA TGGCACTAAA GCGGCTGTCG TTAATGTGCA
1701 GAGTACCAAT GGAGCTCTCA CTATATCCGG AACTTTAGGA CTTGTGAAAA
1751 ACTCTCAAGA TTGCTGTGAC AACCACGGGA TGTTTAATAA AGATTTACAG
1801 CAAGTTCCGA TTTTAGAACT CAAAGCGACT TCAAATACTG TAACCACTAC
40  1851 GGACTTCAGT CTCGGCACAA ACGGCTATCA GCAATCTCCC TATGGGTATC
1901 AAGGAACTTG GGAGTTTACC ATAGACACGA CAACCCATAC GGTACACAGG
1951 AATTGGAAAA AAACCGGTTA TCTTCCTCAT CCGGAGCGTC TTGCTCCCTT
2001 CATTCTAAT AGCCTATGGG CAAACGTCAT AGATTACGA GCTGTAAGTC
2051 AAGCGTCAGC AGCTGATGGC GAAGATGTCC CTGGGAAGCA ACTGAGCATC
45  2101 ACAGGAATTA CAAATTTCTT CCATGCGAAT CATACCGGTG ATGCACGCAG
2151 CTACCGCCAT ATGGGTGGAG GCTACCTCAT CAATACCTAC ACACGCATCA
2201 CTCCAGATGC TCGGTTAAGT CTAGGTTTGG GACAGCTGTT TACAAAATCT
2251 AAGGATTACC TCGTAGGTCA CGGTCAATCT AACGTTTATT TCGCTACAGT
2301 ATACTCTAAC ATCACCAGT CTCTGTTTGG ATCATCGAGA TTCTTCTCAG
50  2351 GAGGCACTTC TCGAGTTACC TATAGCCGTA GCAATGAGAA AGTAAAGACT
2401 TCATATACAA AATTGCCCTAA AGGGCGCTGC TCTTGAGTA ACAATTGCTG
2451 GTTAGGAGAA CTCGAAGGGA ACCTTCCCAT CACTCTCTCT TCTCGCATCT
2501 TAAACCTCAA GCAGATCATT CCCTTTGTAA AAGCTGAAGT TGCTTACCGG
2551 ACTCATGGGG GCATCCAAGA AAATACCCCC GAGGGGAGGA TTTTGGGACA
55  2601 CGGTCATCTA CTCAACGTTG CAGTTCCCGT AGGCGTCCGC TTTGGTAAAA
2651 ATTCTCATAA TCGACCAAGT TTTTACACTA TAATCGTAGC CTATGCTCCT
2701 GATGTCTATC GTCACAATCC TGATTGCGAT ACGACATTAC CTATTAATGG
2751 AGCTACGTGG ACCTCTATAG GGAATAATCT AACCAGAAGT ACTTTGCTAG
2801 TACAAGCATC CAGCCATACT TCAGTAAATG ATGTTCTAGA GATCTTCGGG
2851 CACTGTGGAT GTGATATTCG CAGAACCTCC CGTCAATATA CTCTAGATAT
60  2901 AGGAAGCAAA TTACGATTTT AA

```

The PSORT algorithm predicts an outer membrane location (0.917).



The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 2A, and also as a GST-fusion. Both proteins were used to immunise mice, whose sera were used in a Western blot (Figure 2B) and for FACS analysis (Figure 2C).

The cp6736 protein was also identified in the 2D-PAGE experiment (Cpn0453) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6736 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 3

The following *C.pneumoniae* protein (PID 4376751) was expressed <SEQ ID 5; cp6751>:

```

10      1  MRFFCFGMLL PFTFVLANEG LQLPLETYIT LSPEYQAAPO VGFTHNQNDQ
      51  LAIVGNHNDF ILDYKYYSRN GGALTCKNLL ISENIGNVFF EKNVCPNSGG
     101  AIYAAQNCTI SKNQNYAFTT NLVSDNPTAT AGSLLGGALF AINCSITNNL
     151  GQGTFFVDNLA LNKGGALYTE TNLSIKDNKG PIIKQNRAL NSDSLGGGIY
     201  SGNSLNIEGN SGAIQTTSNS SGSGGGIFST QTLTISSNKK LIBISENSAF
     251  ANNYGSNFNP GGGGLTTTFC TILNNREGVL FNNNQSQSNG GAIHAKSIII
     301  KENGPVYFLN NTATRGGALL NLSAGSGNGS FILSADNGDI IFNNNTASKH
     351  ALNPPYRNAI HSTPNMNLQI GARPGYRVLF YDPIEHELPS SFPILNFET
     401  GHTGTVLFSG EHVHQNFIDE MNFFSYLRNT SELRQGVLA V EDGAGLACYK
     451  FFQRGGTLIL GQGA VITTAG TIPTPSSTPT TVGSTITL NH IAIDLPSILS
     501  FQAQAPKIWI YPTKTGSTYT EDSNPTITIS GTLTLRNSNN EDPYDSL DLS
     551  HSLEKVP LLY IVDVAAQKIN SSQDLDLSTLN SGEHYGYQGI WSTYWVETTT
     601  ITNPTSLLGA NTKHKLLYAN WSPLGYRPHP ERRGEFITNA LWQSAYTALA
     651  GLHSLSSWDE EKGHAA SLOG IGLLVHQKDK NGFKGFRSHM TGYSATTEAT
     701  SSQSPNFSLG FAQFFSKAKE HESQNSTSSH HYFSGMCIEN TLFKEWIRLS
     751  VSLAYMFTSE HTHMYQGLL EGN SQGSFHN HTLAGALSCV FLPQPHGESL
     801  QIYPFITALA IRGNLAAFQE SGDHAREFSL HRPLTDVSLP VGIRASWKNH
     851  HRVPLVWLTE ISYRSTLYRQ DPELHSKLLI SQGTWTTQAT PVTYNALGIK
     901  VKNTMQVFPK VTLSLDYSAD ISSSTLSHYL NVASRMRF*

```

A predicted signal peptide is highlighted.

30 The cp6751 nucleotide sequence <SEQ ID 6> is:

```

      1  ATGCGCTTTT TTTGCTTCGG AATGTTGCTT CCTTTTACTT TTGTATTGGC
     51  TAATGAAGGT CTCCAAC TTC TTTGGAGAC CTATATTACA TTAAGTCCTG
    101  AATATCAAGC AGCCCTCAA GTAGGTTTA CTCATAACCA AAATCAAGAT
    151  CTCGCAATTG TCGGGATCA CAATGATTTT ATCTTGGA CT ATAAGTACTA
    201  TCGGTCGAAT GGAGGTGCTC TTACCTGTAA GAATCTTCTG ATCTCTGAAA
    251  ATATAGGGAA TGCTTCTTTT GAGAAGAATG TCTGTCCCAA TTCTGGCGGG
    301  GCAATTTATG CTGCTCAAAA TTGCACGATC TCCAAGAATC AGAACTATGC
    351  ATTTACTACA AACTTGGTCT CTGACAATCC TACAGCCACT GCGGGATCAC
    401  TATTGGGTGG AGCTCTCTTT GCCATAAATT GCTCTATTAC TAATAACCTA
    451  GGACAGGGAA CTTTCGTTGA CAATCTCGCT TTAAATAAGG GGGGTGCCCT
    501  CTATACTGAG ACGAACTTAT CTATTAAAGA CAATAAAGGC CCGATCATAA
    551  TCAAGCAGAA TCGGGCACTA AATTCGGACA GTTTAGGAGG AGGGATTAT
    601  AGTGGGAACT CTCTAAATAT AGAGGGAAT TCTGGAGCTA TACAGATCAC
    651  AAGCAACTCT TCAGGATCTG GGGGAGGCAT ATTTTCTACC CAAACTCA
    701  CGATCTCCTC GAATAAAAAA CTCATAGAAA TCAGTGAAAA TTCCGCGTTC
    751  GCAAATAACT ATGGATCGAA CTTCAATCCA GGAGGAGGAG GTCTTACTAC
    801  CACCTTTTGC ACGATATTGA ACAACCGAGA AGGGGTACTC TTTAACAATA
    851  ACCAAAGCCA GAGCAACGGT GGAGCCATTC ATGCGAAATC TATCATTATC
    901  AAAGAAAATG GTCCTGTATA CTTTTTAAAT AACACTGCAA CTCGGGGAGG
    951  GGCTCTCCTC AACTTATCAG CAGGTTCTGG AAACGGAAGC TTCATCTTAT
   1001  CTGCAGATAA TGGAGATATT ATCTTTAACA ATAATACGGC CTCCAAGCAT
   1051  GCCCTCAATC CTCCATACAG AAACGCCATT CACTCGACTC CTAATATGAA
   1101  TCTGCAATAA GGAGCCGCTC CCGCTATCG AGTGCTGTTC TATGATCCCA
   1151  TAGAACATGA GCTCCCTTCC TCCTTCCCA TACTCTTTAA TTTCGAAACC
   1201  GGTACATACG GTACAGTTTT ATTTTCAGGG GAACATGTAC ACCAGAACTT

```

5  
10  
15  
20  
25  
30

```

1251 TACCGATGAA ATGAATTTCT TTTCTATT TT AAGGAACACT TCGGAAC TAC
1301 GTC AAGGAGT CCTTGCTGTT GAAGATGGTG CGGGGCTGGC CTGCTATAAG
1351 TTCTTCCAAC GAGGAGGCAC TCTACTTCTA GGTC AAGGTG CGGTGATCAC
1401 GACAGCAGGA ACGATTTCCCA CACCATCCTC AACACCAACG ACAGTAGGAA
1451 GTACTATAAC TTTAAATCAC ATTGCCATTG ACCTTCCTTC TATTCTTTCT
1501 TTTCAAGCTC AGGCTCCAAA AATTTGGATT TACCCACAA AAACAGGATC
1551 TACCTATACT GAAGATTCCA ACCCGACAAT CACAATCTCA GGAAC TCTCA
1601 CCTTACGCAA CAGCAACAAC GAAGATCCCT ACGATAGTCT GGATCTCTCG
1651 CACTCTCTTG AGAAAGTTCC CCTTCTTTAT ATTGTCGATG TCGCTGCACA
1701 AAAAAATTAAC TCTTCGCAAC TGGATCTATC CACATTAAAT TCTGGCGAAC
1751 ACTATGGGTA TCAAGGCATC TGGTCGACCT ATTGGGTAGA AACTACAACA
1801 ATCACGAACC CTACATCTCT ACTAGGCGCG AATACAAAAC ACAAGCTGCT
1851 CTATGCAAAC TGGTCTCCTC TAGGCTACCG TCCTCATCCC GAACGTCGAG
1901 GAGAAATTCAT TACGAATGCC TTGTGGCAAT CGGCATATAC GGCTCTTGCA
1951 GGA CTCCACT CCCTCTCCTC CTGGGATGAA GAGAAGGGTC ATGCAGCTTC
2001 CCTACAAGGC ATTGGTCTTC TGGTTCATCA AAAAGACAAA AACGGTTTTC
2051 AGGGATTTCG TAGTCATATG ACAGGTTATA GTGCTACCAC CGAAGCAACC
2101 TCTTCTCAAA GTCCGAATTT CTCTTTAGGA TTTGCTCAGT TCTTCTCCAA
2151 AGCTAAAGAA CATGAATCTC AAAATAGCAC GTCCTCTCAC CACTATTTCT
2201 CTGGAATGTG CATAGAAAAT ACTCTCTTCA AAGAGTGGAT ACGTCTATCT
2251 GTGTCTCTTG CTTATATGTT TACCTCGGAA CATACCCATA CAATGTATCA
2301 GGGTCTCCTG GAAGGGAAC CTACAGGATC TTTCCACAAC CATACCTTAG
2351 CAGGGGCTCT CTCCTGTGTT TTCTTACCTC AACCTCACGG CGAGTCCCTG
2401 CAGATCTATC CTTTATTAC TGCCTTAGCC ATCCGAGGAA ATCTTCTGTC
2451 GTTTCAAGAA TCTGGAGACC ATGCTCGGGA ATTTTCCCTA CACCGCCCCC
2501 TAACGGACGT CTCCCTCCCT GTAGGAATCC GCGCTTCTTG GAAGAACCAC
2551 CACCGAGTTC CCCTAGTCTG GCTCACAGAA ATTTCTTATC GCTCTACTCT
2601 CTATAGGCAA GATCCTGAAC TCCACTCGAA ATTACTGATT AGCCAAGGTA
2651 CGTGGACGAC GCAGGCCACT CCTGTGACCT ACAATGCTTT AGGGATCAAA
2701 GTGAAAAATA CCATGCAGGT GTTTCCTAAA GTCACTCTCT CCTTAGATTA
2751 CTCTGCGGAT ATTTCTTCTT CCACGCTGAG TCACTACTTA AACGTGGCGA
2801 GTAGAATGAG ATTTTAA

```

The PSORT algorithm predicts an outer membrane location (0.923).

35 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 3A, and also in his-tagged form. The GST-fusion recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 3B) and for FACS analysis (Figure 3C).

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

40 These experiments show that cp6751 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 4

The following *C.pneumoniae* protein (PID 4376752) was expressed <SEQ ID 7; cp6752>:

45  
50  
55

```

1 MFGMTPAVYS LQTDLSLEKFA LERDEEFRTS FPLLDLSLSTL TGFSPITTFV
51 GNRHNSSQDI VLSNYKSIDN ILLLWTSAGG AVSCNNFLLS NVEDHAFFSK
101 NLAIGTGGAI ACQGACTITK NRGPLIFFSN RGLNNASTGG ETRGGAIACN
151 GDFITISQNG TFYFVNNSVN NWGGALSTNG HCRIQSNRAP LLFFNNTAPS
201 GGGALRSENT TISDNTRPIY FKNNGGNNGG AIQTSVTVAI KNNSGSVIFN
251 NNTALSGSIN SGNGSGGAIY TTNLSIDNPN GTILFNNNYC IRDGGAICTQ
301 FLTIKNSGHV YFTNNQGNWG GALMLLQDST CLLFAEQGNI AFQNNNEVFLT
351 TFGRYNAIHC TPNSNLQLGA NKGYTATFFD PIEHQHPTTN PLIFNPANAH
401 QGTILFSSAY IPEASDYENN FISSSKNTSE LRNGVLSIED RAGWQFYKFT
451 QKGGLILKLGH AASIATTANS ETPSTSVGSQ VIINNLAIDL PSILAKGKAP
501 TLWIRPLQSS APFTEDNNPT ITLSGPLTLL NEENRDPYDS IDLSEPLQNI
551 HLLSLSDVTA RHINTDNFHP ESLNATEHYG YQGIWSPYVW ETITTTNNAS
601 IETANTLYRA LYANWTPLGY KVNPEYQDGL ATTPWLQSFH TMFSLRLSYN
651 RTGDSDIERP FLEIQGIADG LFVHQNSIPG APGFRIQSTG YSLQASSETS

```

-45-

701 LHQKISLGFA QFFTRTKKEIG SSNNVSAHNT VSSLYVELPW FQEAFASTSV  
 751 LAYGYGDHHL HSLHPSHQEQ AEGTCYSHLT AAAIGCSFPW QQKSYLHLSLSP  
 801 FVQAIAIRSH QTAFFEEIGDN PRKFVSQKPF YNLTLPLGIQ GKWQSKFHVFP  
 851 TEWTTLELSYQ PVLYQONPQI GVTLLASGGS WDILGHNYVR NALGYKVHNO  
 901 TALFRSLDLF LDYQGSVSSS TSTHHLQAGS TLKF\*

The cp6752 nucleotide sequence <SEQ ID 8> is:

1 ATGTTTCGGGA TGAATCCTGC AGTGTATAGT TTACAAACGG ACTCCCTTGA  
 51 AAAGTTTGCT TTAGAGAGGG ATGAAGAGTT TCGTACGAGC TTTCTCTCTCT  
 101 TAGACTCTCT CTCCACTCTT ACAGGATTTT CTCCAATAAC TACGTTTGTT  
 151 GGAAATAGAC ATAATTCCTC TCAAGACATT GTACTTTCTA ACTACAAGTC  
 201 TATTGATAAC ATCCTTCTTC TTTGGACATC GGCTGGGGGA GCTGTGTCCT  
 251 GTAATAATTT CTTATTATCA AATGTTGAAG ACCATGCCTT CTTCAAGTAA  
 301 AATCTCGCGA TTGGGACTGG AGGCGCGATT GCTTGCCAGG GAGCCTGCAC  
 351 AATCAGCAAG AATAGAGGAC CCCTTATTTT TTTCAAGCAAT CGAGGTCTTA  
 401 ACAATGCGAG TACAGGAGGA GAAACTCGTG GGGGTGCGAT TGCCTGTAAT  
 451 GGAGACTTCA CGATTTCTCA AAATCAAGGG ACTTTCTACT TTGTCAACAA  
 501 TTCCGTC AAC TGGGGGAG GAGCCCTCTC CACCAATGGA CACTGCCGCA  
 551 TCCAAAGCAA CAGGGCACCT CTACTCTTTT TTAACAATAC AGCCCCTAGT  
 601 GGAGGGGGTG CGCTTCGTAG TGAAAATACA ACGATCTCTG ATAACACGCG  
 20 TCCATTTTAT TTTAAGAAC AACTGTGGGA CAATGGCGGG GCCATTCAAA  
 701 CAAGCGTTAC TGTGCGATA AAAAATAACT CCGGGTCGGT GATTTTCAAT  
 751 AACACACAG CGTTATCTGG TTCGATAAAT TCAGGAAATG GTTCAGGAGG  
 801 GCGGATTTAT ACAACAAACC TATCCATAGA CGATAACCTT GGAACATATC  
 851 TTTTCAATAA TAACTACTGC ATTCGCGATG GCGGAGCTAT CTGTACACAA  
 25 TTTTGTACAA TCAAAAATAG TGGCCACGTA TATTTACCA ACAATCAAGG  
 951 AAAGTGGGA GGTGCTCTTA TGCTCCTACA GGACAGCACC TGCCCTACTCT  
 TCGCGGAACA AGGAAATATC GCATTTCAAA ATAATGAGGT TTTCTCACC  
 1001 ACATTTGGTA GATACAACGC CATACATTGT ACACCAATA GCAACTTACA  
 1051 ACTTGGAGCT AATAAGGGGT ATACGACTGC TTTTTTTGAT CCTATAGAAC  
 1101 ACCAACATCC AACTACAAAT CCTCTAATCT TTAATCCCAA TCGGAACCAT  
 1151 CAGGGAACGA TCTTATTTTC TTCAGCCTAT ATCCCAGAAG CTTCTGACTA  
 1201 CGAAAATAAT TTCATTAGCA GCTCGAAAAA TACCTCTGAA CTTGCAATG  
 1251 GTGTCCTCTC TATCGAGGAT CGTGCGGGAT GGCAATTCTA TAAGTTCACT  
 1301 CAAAAAGGAG GTATCCTTAA ATTAGGGCAT GCGGCGAGTA TTGCAACAAC  
 1351 TGCCAACTCT GAGACTCCAT CAACTAGTGT AGGCTCCCAG GTCATCATTA  
 1401 ATAACCTTGC GATTAACCTC CCCTCGATCT TAGCAAAAGG AAAAGCTCCT  
 1451 ACCTTGTGGA TCCGTCCTCT ACAATCTAGT GCTCCTTTCA CAGAGGACAA  
 1501 TAACCCTACA ATTACTTTAT CAGGTCCTCT GACACTCTTA AATGAGGAAA  
 1551 ACCGCGATCC CTACGACAGT ATAGATCTCT CTGAGCCTTT ACAAAACATT  
 1601 CATCTTCTTT CTTTATCGGA TGTAACAGCA CGTCATATCA ATACCGATAA  
 40 CTTTCATCCT GAAAGCTTAA ATGCGACTGA GCATTACGGT TATCAAGGCA  
 1701 TCTGGTCTCC TTATTGGGTA GAGACGATAA CAACAACAAA TAACGCTTCT  
 1751 ATAGAGACGG CAAACACCCCT CTACAGAGCT CTGTATGCCA ATTGGACTCC  
 1801 CTTAGGATAT AAGGTCAATC CTGAATACCA AGGAGATCTT GCTACGACTC  
 1851 CCCTATGGCA ATCCTTTCAT ACTATGTTCT CTCTATTAAG AAGTTATAAT  
 1901 CGAACTGGTG ATTCTGATAT CGAGAGGCCCT TTCTTAGAAA TTCAAGGGAT  
 1951 TGCCGACGGC CTCTTTGTTT ATCAAAATAG CATCCCCGGG GCTCCAGGAT  
 2001 TCCGTATCCA ATCTACAGGG TATTCCTTAC AAGCATCCTC CGAAACTTCT  
 2051 TTACATCAGA AAATCTCCTT AGGTTTTGCA CAGTTCTTCA CCCGCACTAA  
 2101 AGAAATCGGA TCAAGCAACA ACGTCTCGGC TCACAATACA GTCTCTTAC  
 2151 TTTATGTTGA GCTTCCGTGG TTCCAAGAGG CCTTTGCAAC ATCCACAGTG  
 2201 TTAGCGTATG GCTATGGGGA CCATCACCTC CACAGCCTAC ATCCCTCACA  
 2251 TCAAGAACAG GCAGAAGGGA CGTGTATAG CCATACATTA GCAGCAGTA  
 2301 TCGGCTGTTC TTTCCCTTGG CAACAGAAAT CCTATCTTCA CCTCAGCCCG  
 2351 TTCGTTCAAG CAATTGCAAT ACGTTCTCAC CAAACAGCGT TCGAAGAGAT  
 2401 TGGTGACAAT CCCCAGAAAGT TTGTCTCTCA AAAGCCTTTC TATAATCTGA  
 2451 CCTTACCTCT AGGAATCCAA GGAAAATGGC AGTCAAAAT CCACGTACCT  
 2501 ACAGAATGGA CTCTAGAACT TTCTTACCAA CCGGTACTCT ATCAACAAA  
 2551 TCCCCAAATC GGTGTACGC TACTTGCGAG CGGAGGTTCC TGGGATATCC  
 2601 TAGGCCATAA CTATGTTTCGC AATGCTTTAG GGTACAAAGT CCACAATCAA  
 2651 ACTGCGCTCT TCCGTTCTCT CGATCTATTC TTGGATTACC AAGGATCGGT  
 2701 CTCCTCCTCG ACATCTACGC ACCATCTCCA AGCAGGAAGT ACCTTAAAT  
 2751 TCTAA

The PSORT algorithm predicts a cytoplasmic location (0.138).

The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 4A, and also as a GST-fusion. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (4B) and the his-tagged protein was used for FACS analysis (4C).

The cp6752 protein was also identified in the 2D-PAGE experiment (Cpn0467).

- 5 These experiments show that cp6752 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 5

The following *C.pneumoniae* protein (PID 4376850) was expressed <SEQ ID 9; cp6850>:

10 1 MKKAVLIAAM FCGVVSLSSC CRIVDCCFED PCAPSSCNPC EVIRKKERSC  
51 GGNACGSYVP SCSNPCGSTE CNSQSPQVKG CTSPDGRCKQ \*

A predicted signal peptide is highlighted.

The cp6850 nucleotide sequence <SEQ ID 10> is:

15 1 ATGAAGAAAG CTGTTTAAAT TGCTGCAATG TTTTGTGGAG TAGTTAGCTT  
51 AAGTAGCTGC TGCCGCATTG TAGATTGTTG TTTTGAGGAT CCTTGCGCAC  
101 CCTCTTCTTG CAATCCTTGT GAAGTAATAA GAAAAAAGA AAGATCTTGC  
151 GCGGTAATG CTTGTGGGTC CTACGTTCTT TCTGTGTTCTA ATCCATGTGG  
201 TTCAACAGAG TGTAACTCTC AAAGCCCACA AGTTAAAGGT TGTACATCAC  
251 CTGATGGCAG ATGCAAACAG TAA

The PSORT algorithm predicts an inner membrane location (0.329).

- 20 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 5A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 5B) and for FACS analysis (Figure 5B). A his-tagged protein was also expressed.

These experiments show that cp6850 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### 25 Example 6

The following *C.pneumoniae* protein (PID 4376900) was expressed <SEQ ID 11; cp6900>:

30 1 MKIKFSWKVN FLICLLAVGL IFFGCSRVRK EVLVGRDATW FPKQFGIYTS  
51 DTNAFLNDLV SEINYKENLN INIVNQDWVH LFENLDDKKT QGAFTSVLPT  
101 LEMLEHYQFS DPILLTGPVL VVAQDSPYQS IEDLKGRLLG VYKFDSSVLV  
151 AQNIPDAVIS LYQHVPIALE ALTSNCDYDAL LAPVIEVTAL IETAYKGRLLK  
201 IISKPLNADG LRLAILKGTN GDLLEGFNAG LVKTRRSRGKY DAIKQRYRLP

The cp6900 nucleotide sequence <SEQ ID 12> is:

35 1 GTGAAGATAA AATTTTCTTG GAAGGTAAAT TTTTAAATAT GTTTACTGGC  
51 TGTGGGACTG ATCTTTTTCG GGTGCTCTCG AGTAAAAAGA GAAGTTCTCG  
101 TAGGTCGTGA TGCCACCTGG TTTCCAAAAC AATTCGGCAT TTATACATCC  
151 GATACCAACG CATTTTAAA CGATCTTGTT TCTGAGATTA ACTATAAAGA  
201 GAATCTAAAT ATTAATATTTG TAAATCAAGA TTGGGTGCAT CTCCTTGGAGA  
251 ATTTAGATGA TAAAAAGACC CAAGGAGCAT TTACATCTGT ATTGCCTACT  
301 CTTGAGATGC TCGAACACTA TCAATTTTCT GATCCCATTT TACTCACAGG  
40 351 TCCTGTCTCTT TCGCTCGCTC AAGACTCTCC TTACCAATCT ATAGAGGATC  
401 TTAAAGGTCG TCTTATTGGA GTGTATAAGT TTGACTCTTC AGTTCCTGTGA  
451 GCTCAAAATA TCCCTGACGC TGTGATTAGC CTCTACCAAC ATGTTCCAAT  
501 AGCATTGGAA GCCTTAACAT CGAATTGTTA CGACGCTCTT CTAGCTCCTG  
551 TAATTGAAGT GACCGCGCTA ATAGAAACAG CATATAAAGG AAGACTGAAA  
45 601 ATTATTTCAA AACCTTAAAC CGCAGATGGT TTGCGGCTTG CAATACTGAA

651 AGGGACAAAC GGAGATTTGC TTGAAGGGTT TAACGCAGGA CTTGTGAAAA  
 701 CACGACGCTC AGGAAAATAC GATGCTATAA AACAGCGGTA TCGTCTTCCC  
 751 TAA

The PSORT algorithm predicts an inner membrane location (0.452).

- 5 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 6A. The recombinant protein was used to immunise mice, whose sera were used for FACS analysis (Figure 6B). A his-tagged protein was also expressed.

The cp6900 protein was also identified in the 2D-PAGE experiment (Cpn0604).

- 10 These experiments show that cp6900 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 7

The following *C.pneumoniae* protein (PID 4377033) was expressed <SEQ ID 13; cp7033>:

15  
 20  
 25

```

1  MVNPIGPGPI DETERTPPAD LSAQGLEASA ANKSAEAQRI AGAEAKPKES
51  KTDSVERWSI LRSVAVNALMS LADKLGIIASS NSSSSTSRSA DVDSTTATAP
101 TPPPPTFDDY KTQAQTAYDT IFTSTSLADI QAALVSLQDA VTNIKDТАAT
151 DEETAIAAEW ETKNADAVKV GAQITELAKY ASDNQAILDS LGKLTSFDDL
201 QAALLQSVAN NNKAAELLKE MQDNPVVPGK TPAIAQSLVD QTDATATQIE
251 KDGNAIRDAY FAGQNASGAV ENAKSNNSIS NIDSAKAAIA TAKTQIAEAQ
301 KKFPDSPILQ EAEQMVIQAE KDLKNIKPAD GSDVPNPGTT VGGSKQQGSS
351 IGSIRVSMIL DDAENETASI LMSGFRQMIH MFNTENPDSQ AAQOELAAQA
401 RAAKAAGDDS AAAALADAQK ALEAALGKAG QQQGIILNALG QIASAAVVSА
451 GVPPAAASSI GSSVKQLYKT SKSTGSDYKT QISAGYDAYK SINDAYGRAR
501 NDATRDVINN VSTPALTRSV PRARTEARGP EKTDQALARV ISGNSRTLGD
551 VYSQVSALQS VMQIIQSNPQ ANNEEIRQKL TSAVTKPPQF GYPYVQLSND
601 STQKFIKLE SLFAEGSRТА AEIKALSFET NSLFIQQVLV NIGSLYSGYL
651 Q*
```

The cp7033 nucleotide sequence <SEQ ID 14> is:

30  
 35  
 40  
 45  
 50  
 55

```

1  ATGGTTAATC CTATTGGTCC AGGTCCTATA GACGAAACAG AACGCACACC
51  TCCCGCAGAT CTTTCTGCTC AAGGATTGGA GCGAGTGCA GCAAATAAGA
101 GTGCGGAAGC TCAAAGAATA GCAGGTGCGG AAGCTAAGCC TAAAGAATCT
151 AAGACCGATT CTGTAGAGCG ATGGAGCATC TTGCGTTC TG CAGTGAATGC
201 TCTCATGAGT CTGGCAGATA AGCTGGGTAT TGCTTCTAGT AACAGCTCGT
251 CTTCTACTAG CAGATCTGCA GACGTGGACT CAACGACAGC GACCGCACCT
301 ACGCCTCCTC CACCCACGTT TGATGATTAT AAGACTCAAG CGCAAACAGC
35  351 TTACGATACT ATCTTTACCT CAACATCACT AGCTGACATA CAGGCTGCCT
401 TGGTGAGCCT CCAGGATGCT GTCACТАATA TAAAGGATAC AGCGGCTACT
451 GATTAGGAAA CCGCAATCGC TGCGGAGTGG GAAACTAAGA ATGCCGATGC
501 AGTTAAAGTT GGCGCGCAAA TTACAGAATT AGCGAAATAT GCTTCGGATA
551 ACCAAGCGAT TCTTGACTCT TTAGGTAAAC TGACTTCCTT CGACCTCTTA
40  601 CAGGCTGCTC TTCTCCAATC TGТАGCAAAC AATAACAAAG CAGCTGAGCT
651 TCTTAAAGAG ATGCAAGATA ACCCAGTAGT CCCAGGGAАA ACGCCTGCAA
701 TTGCTCAATC TTTAGTTGAT CAGACAGATG CTACAGCGAC ACAGATAGAG
751 AAAGATGGAA ATGCGATTAG GGATGCATAT TTTGCAGGAC AGAACGCTAG
801 TGGAGCTGTA GAAAATGCTA AATCTAATAA CAGТАAAGC AACATAGATT
45  851 CAGCTAAAGC AGCAATCGCT ACTGCTAAGA CACAAATAGC TGAAGCTCAG
901 AAAAAAGTTC CCGACTCTCC AATTCTTCAA GAAGCGGAAC AAATGGTAAT
951 ACAGGCTGAG AAAGATCTTA AAAATATCAA ACCTGCAGAT GGTTC TGATG
1001 TTCCAAATCC AGGAACTACA GTTGAGGGCT CCAAGCAACA AGGAAGTAGT
1051 ATTGGTAGTA TTCGTGTTTC CATGCTGTTA GATGATGCTG AAAATGAGAC
50  1101 CGGTTCCATT TTGATGCTG GGTTCGTCА GATGATTCAC ATGTTCAATA
1151 CGGAAATCC TGATTCTCAA GCTGCCCAAC AGGAGCTCGC AGCACAAGCT
1201 AGAGCAGCGA AAGCCGCTGG AGATGACAGT GCTGCTGCAG CGCTGGCAGA
1251 TGCTCAGAAA GCTTTAGAAG CGGCTCTAGG TAAAGCTGGG CAACAACAGG
1301 GCATACTCAA TGCTTTAGGA CAGATCGCTT CTGCTGCTGT TGTGAGCGCA
55  1351 GGAGTTCCTC CCGCTGCAGC AAGTTCТАA GGGTCATCTG TAAAACAGCT
1401 TTACAAGACC TCAAAATCTA CAGGTTCTGA TTATAAAACA CAGATATCAG
```

1451 CAGGTTATGA TGCTTACAAA TCCATCAATG ATGCCTATGG TAGGGCACGA  
 1501 AATGATGCGA CTCGTGATGT GATAACAAT GTAAGTACCC CCGCTCTCAC  
 1551 ACGATCCGTT CCTAGAGCAC GAACAGAAGC TCGAGGACCA GAAAAACAG  
 5 1601 ATCAAGCCCT CGCTAGGGTG ATTTCTGGCA ATAGCAGAAC TCTTGAGAT  
 1651 GTCTATAGTC AAGTTTCGGC ACTACAATCT GTAATGCAGA TCATCCAGTC  
 1701 GAATCCTCAA GCGAATAATG AGGAGATCAG ACAAAGCTT ACATCGGCAG  
 1751 TGACAAAGCC TCCACAGTTT GGCTATCCTT ATGTGCAACT TTCTAATGAC  
 1801 TCTACACAGA AGTTCATAGC TAAATTAGAA AGTTTGTGTTG CTGAAGGATC  
 1851 TAGGACAGCA GCTGAAATAA AAGCACTTTC CTTTGAAACG AACTCCTTGT  
 10 1901 TTATTCAGCA GGTGCTGGTC AATATCGGCT CTCTATATTC TGGTTATCTC  
 1951 CAATAA

The PSORT algorithm predicts a cytoplasmic location (0.272).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 7A. A  
 his-tagged protein was also expressed. The recombinant proteins were used to immunise mice, whose  
 15 sera were used for FACS (Figure 7B) and Western blot (7C) analyses.

The cp7033 protein was also identified in the 2D-PAGE experiment (Cpn0728) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7033 a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## 20 Example 8

The following *C.pneumoniae* protein (PID 6172321) was expressed <SEQ ID 15; cp0017>:

1 MGIKGTGIIV WVDDATAKTK NATLTWTKTG YKPNPERQGP LVPNSLWGSF  
 51 VDVRSIQSLM DRSTSSLSSS TNLWVSGIAD FLHEDQKGNQ RSYRHSSAGY  
 101 ALGGGFFTAS ENFFNFAFCQ LFGYDKDHLV AKNHTHVYAG AMSYRHLGES  
 25 151 KTLAKILSGN SDSLPFVFNA RFAYGHTDNN MTTKYTGYSF VKGSWGNDAF  
 201 GIECGGAIPV VASGRRSWVD THTPFLNLEM IYAHQNDPKE NGTEGRSFQS  
 251 EDLFNLAVPV GIKFEKFSK STYDLSTIAYV PDVIRNDPGC TTTLMVSGDS  
 301 WSTCGTSLSR QALLVRAGNH HAFASNFEVF SQFEVELRGS SRSYATDLGG  
 351 RFGF\*

30 The cp0017 nucleotide sequence <SEQ ID 16> is:

1 ATGGGTATCA AGGGAACCTG AATAATTGTT TGGGTCGACG ATGCAACTGC  
 51 AAAAAACAAA AATGCTACCT TAACTTGGAC TAAAACAGGA TACAAGCCGA  
 101 ATCCAGAACG TCAGGGACCT TTGGTTCCTA ATAGCCTGTG GGGTTCCTTT  
 151 GTCGATGTCC GCTCCATTCA GAGCCTCATG GACCGGAGCA CAAGTTCGTT  
 35 201 ATCTTCGTCA ACAAATTGTT GGGTATCAGG AATCGCGGAC TTTTTCGATG  
 251 AAGATCAGAA AGGAAACCAA CGTAGTTATC GTCATTCTAG CGCGGGTTAT  
 301 GCATTAGGAG GAGGATTCTT CACGGCTTCT GAAAATTCTT TTAATTTTGC  
 351 TTTTTGTGAG CTTTTTGGCT ACGACAAGGA CCATCTTGTG GCTAAGAACC  
 401 ATACCCATGT ATATGCAGG GCAATGAGTT ACCGACACCT CGGAGAGTCT  
 451 AAGACCCTCG CTAAGATTTT GTCAGGAAAT TCTGACTCCC TACCTTTTGT  
 501 CTTCAATGCT CGGTTTGCTT ATGGCCATAC CGACAATAAC ATGACCACAA  
 551 AGTACACTGG CTATTCTCCT GTTAAGGGAA GCTGGGGGAA TGATGCGCTT  
 601 GGTATAGAAT GTGGAGGAGC TATCCCGGTA GTTGCTTCAG GACGTCGGTC  
 651 TTGGGTGGAT ACCCACACGC CATTCTTAAA CCTAGAGATG ATCTATGCAC  
 45 701 ATCAGAATGA CTTTAAGGAA AACGGCACAG AAGGCCGTTT TTTCCAAAGT  
 751 GAAGACCTCT TCAATCTAGC GGTTCCTGTA GGGATAAAAT TTGAGAAATT  
 801 CTCCGATAAG TCTACGTATG ATCTCTCCAT AGCTTACGTT CCCGATGTGA  
 851 TTCGTAATGA TCCAGGCTGC ACGACAATC TTATGGTTTC TGGGGATTCT  
 901 TGGTCGACAT TGGGTACAAG CTTGTCTAGA CAAGCTCTTC TTGTACGTGC  
 50 951 TGGAAATCAT CATGCCTTTG CTTCAAACCT TGAAGTTTTC AGTCAGTTTG  
 1001 AAGTCGAGTT GCGAGGTTCT TCTCGTAGCT ATGCTATCGA TCTTGGAGGA  
 1051 AGATTCCGAT TTAA

This sequence is frame-shifted with respect to cp0016.

The PSORT algorithm predicts a cytoplasmic location (0.075).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 8A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 8B) and for FACS analysis (Figure 8C). A his-tagged protein was also expressed.

- 5 This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp0017 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 9

- 10 The following *C.pneumoniae* protein (PID 6172315) was expressed <SEQ ID 17; cp0014>:

```

1  MKSSFPKFVF STFAIFPLSM IATETVLDSS ASFDGNKNGN FSVRESQEDA
51  GTTYLFRGNV TLENIPGTGT AITKSCFNNT KGDLTFTGNG NSLLFQTVDA
101 GTVAGAAVNS SVVDKSTTFI GFSSLSFIAS PGSSIITGKG AVSCSTGSLS
151 LTKMSVCSSA KTFQRIMAVL SPQKLFH*
```

- 15 The cp0014 nucleotide sequence <SEQ ID 18> is:

```

1  ATGAAGTCTT CTTTCCCCAA GTTTGTATTT TCTACATTG CTATTTTCCC
51  TTTGTCTATG ATTGCTACCG AGACAGTTT GGATTCAAGT GCGAGTTTCG
101 ATGGGAATAA AAATGGTAAT TTTTCAGTTC GTGAGAGTCA GGAAGATGCT
151 GGAACACCT ACCTATTTAA GGGAAATGTC ACTCTAGAAA ATATTCTTGG
20  201 AACAGGCACA GCAATCACAA AAAGCTGTTT TAACAACACT AAGGGCGATT
251 TGACTTTTAC AGGTAACGGG AACTCTCTAT TGTTCCAAAC GGTGGATGCA
301 GGGACTGTAG CAGGGGCTGC TGTTAACAGC AGCGTGGTAG ATAAATCTAC
351 CACGTTTATA GGGTTTCTT CGCTATCTT TATTGCGTCT CCTGGAAGTT
401 CGATAACTAC CGGCAAAGGA GCCGTTAGCT GCTCTACGGG TAGCTTGAGT
25  451 TTGACAAAAA TGTCAGTTTG CTCTTCAGCA AAAACTTTTC AACGGATAAT
501 GCGGTGCTA TCACCGCAA AACTCTTCA TTAA
```

This protein is frame-shifted with respect to cp0015.

The PSORT algorithm predicts an inner membrane location (0.047).

- 30 The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 9A. A GST-fusion was also expressed. The recombinant proteins were used to immunise mice, whose sera were used in an immunoassay (Figure 9B) and for FACS analysis (Figure 9C).

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

- 35 These experiments suggest that cp0014 is a useful immunogen. These properties are not evident from the sequence alone.

### Example 10

The following *C.pneumoniae* protein (PID 6172317) was expressed <SEQ ID 19; cp0015>:

```

1  MSALFSENTS SKKGGAIQTS DALTITGNQG EVSFSNTSS DSGAAIFTEA
51  SVTISNNAKV SFIDNKVTGA SSSTTGDMG GAICAYKTST DTKVTLTGNQ
40  101 MLLFSNNTST TAGGAIYVKK LELASGGLTL FSRNSVNGGT APKGGAIAIE
151 DSGELSLSAD SGDIVFLGNT VTSTTPGTNR SSIDLGTS AK MTALRSAAGR
```

-50-

5  
10

```

201 AIYFYDPITT GSSTTVTDVL KVNETPADSA LQYTGNIIFT GEKLSETEAA
251 DSKNLTSKLL QPVTLSGGTL SLKHGVTLOQ QAFTQQADSR LEMDVGTTLLE
301 PADTSTINNL VINISSIDGA KKAKIETKAT SKNLTLSGTI TLLDPTGTFY
351 ENHSLRNPQS YDILELKASG TVTSTAVTPD PIMGEKFHYG YQGTWGPVW
401 GTGASTTATF NWTKTGYIPN PERIGSLVPN SLWNAFIDIS SLHYLMETAN
451 EQLQGDRAFW CAGLSNFFHK DSTKTRRGFR HLSGGYVIGG NLHTCSDKIL
501 SAAFCQLFGR DRDYFVAKNQ GTVYGGTLYY QHNETYISLP CKLRPCSLSY
551 VPTEIPVLFS GNLSYTHTDN DLKTKYTTYYP TVKGSWGND SFALEFGGRAP
601 ICLDESALFE QYMPFMKLQF VYAHQEGFKE QGTEAREFGS SRLVNLA LPI
651 GIRFDKESDC QDATYNLTG YTVDLVRSNP DCTTTLRISG DSWKTFGTNL
701 ARQALVLRAG NHFCFNSNFE AFSQFSFELR GSSRNYNVDL GAKYQF*

```

This sequence is frame-shifted with respect to cp0014.

The cp0015 nucleotide sequence <SEQ ID 20> is:

15  
20  
25  
30  
35  
40  
45  
50  
55

```

1 ATGTCAGCTC TGTTTTCTGA AAATACCTCC TCAAAGAAAG GCGGAGCCAT
51 TCAGACTTCC GATGCCCTTA CCATTACTGG AAACCAAGGG GAAGTCTCTT
101 TTTCTGACAA TACTTCTTCG GATTCTGGAG CTGCAATTTT TACAGAAGCC
151 TCGGTGACTA TTTCTAATAA TGCTAAAGTT TCCTTTATTG ACAATAAGGT
201 CACAGGAGCG AGCTCCTCAA CAACGGGGGA TATGTCAGGA GGTGCTATCT
251 GTGCTTATAA AACTAGTACA GATACTAAGG TCACCCTCAC TGGAAATCAG
301 ATGTTACTCT TCAGCAACAA TACATCGACA ACAGCGGGAG GAGCTATCTA
351 TGTGAAAAAG CTCGAAGTGG CTTCCGGAGG ACTTACCCTA TTCAGTAGAA
401 ATAGTGTCAA TGGAGGTACA GCTCCTAAAG GTGGAGCCAT AGCTATCGAA
451 GATAGTGGGG AATTGAGTTT ATCCGCCGAT AGTGGTGACA TTGTCTTTTT
501 AGGGAATACA GTCACCTCTA CTACTCCTGG GACGAATAGA AGTAGTATCG
551 ACTTAGGAAC GAGTGCAAG ATGACAGCTT TGCCTTCTGC TGCTGGTAGA
601 GCCATCTACT TCTATGATCC CATAACTACA GGATCATCCA CAACAGTTAC
651 AGATGTCTTA AAAGTTAATG AGACTCCGGC AGATTCTGCA CTACAATATA
701 CAGGGAACAT CATCTTCACA GGAGAAAAGT TATCAGAGAC AGAGGCCCGA
751 GATTCTAAAA ATCTTACTTC GAAGCTACTA CAGCCTGTAA CTCTTTCAGG
801 AGGTACTCTA TCTTTAAAA ATGGAGTGAC TCTGCAGACT CAGGCATTCA
851 CTCAACAGGC AGATTCTCGT CTCGAAATGG ACGTAGGAAC TACTCTAGAA
901 CCTGCTGATA CTAGCACCAT AAACAATTTG GTCATTAACA TCAGTTCTAT
951 AGACGGTGCA AAGAAGGCAA AAATAGAAAC CAAAGCTACG TCAAAAAATC
1001 TGACTTTATC TGGAAACCATC ACTTTATTGG ACCCGACGGG CACGTTTTAT
1051 GAAATCATA GTTTAAGAAA TCCTCAGTCC TACGACATCT TAGAGCTCAA
1101 AGCTTCTGGA ACTGTAACAA GCACCGCAGT GACTCCAGAT CCTATAATGG
1151 GTGAGAAATT CCATTACGGC TATCAGGGAA CTTGGGGCCC AATTGTTTGG
1201 GGGACAGGGG CTTCTACGAC TGCAACCTTC AACTGGACTA AACTGGCTA
1251 TATTCTTAAT CCCGAGCGTA TCGGCTCTTT AGTCCCTAAT AGCTTATGGA
1301 ATGCATTTAT AGATATTAGC TCTCTCCATT ATCTTATGGA GACTGCAAAC
1351 GAAGGGTTGC AGGGAGACCG TGCTTTTGG TGTGCTGGAT TATCTAACTT
1401 CTTCCATAAG GATAGTACAA AAACACGACG CGGGTTTCGC CATTGAGTG
1451 GCGGTTATGT CATAGGAGGA AACCTACATA CTGTTCAGA TAAGATTCTT
1501 AGTGCTGCAT TTTGTCAGCT CTTTGAAGA GATAGAGACT ACTTTGTAGC
1551 TAAGAAATCAA GGTACAGTCT ACGGAGGAAC TCTCTATTAC CAGCACAAACG
1601 AAACCTATAT CTCTCTTCTT TGCAAACTAC GGCTTGTTC GTTGTCTTAT
1651 GTTCCTACAG AGATTCCTGT TCTCTTTTCA GGAACCTTA GCTACACCCA
1701 TACGGATAAC GATCTGAAA CCAAGTATAC AACATATCCT ACTGTTAAAG
1751 GAAGCTGGGG GAATGATAGT TTCGCTTTAG AATTCCGGTG AAGAGCTCCG
1801 ATTTGCTTAG ATGAAAGTGC TCTATTGAG CAGTACATGC CCTTCATGAA
1851 ATTGCAGTTT GTCTATGCAC ATCAGGAAGG TTTTAAAGAA CAGGGAACAG
1901 AAGCTCGTGA ATTTGGAAGT AGCCGTCTTG TGAATCTTGC CTTACCTATC
1951 GGGATCCGAT TTGATAAGGA ATCAGACTGC CAAGATGCAA CGTACAATCT
2001 AACTCTTGGT TATACTGTGG ATCTTGTTCG TAGTAACCCC GACTGTACGA
2051 CAACACTGCG AATTAGCGGT GATTCTTGGA AAACCTTCGG TACGAATTTG
2101 GCAAGACAAG CTTTAGTCCT TCGTGACAGG AACCATTTTT GCTTTAACTC
2151 AAATTTTGAA GCCTTTAGCC AATTTCTTT TGAATTGCGT GGGTCATCTC
2201 GCAATTACAA TGTAGACTTA GGAGCAAAAT ACCAATTCTA A

```

The PSORT algorithm predicts a cytoplasmic location (0.274).

60 The protein was expressed in *E. coli* and purified as a GST-fusion product, as shown in Figure 10A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 10B) and for FACS analysis. A his-tagged protein was also expressed.



These experiments show that cp0015 is a useful immunogen. These properties are not evident from the sequence alone.

### Example 11

The following *C.pneumoniae* protein (PID 6172325) was expressed <SEQ ID 21; cp0019>:

```

5      1  LQDSQDYSFV KLSFGAGGTI ITQDASQKPL EVAPSRPHYG YQGHWNVQVI
      51  PGTGTQPSQA NLEWVRTGYL PNPERQGSV PLSLWGSFVD QRAIQEIMVN
     101  SSQILCQERG VWGAGIANFL HRDKINEHGY RHSGVGYLVG VGTHAFSDAT
     151  INAAFCQLFS RDKDYVVSKN HGTSYSGVVF LEDTLEFRSP QGFYTDSSSE
     201  ACCNQVVTID MQLSYSHRNN DMKTKYTTYT EAQGSWANDV FGLEFGATTY
    10  251  YYPNSTFLFD YYSPFLRLQC TYAHQEDFKE TGGEVRHFTS GDLFNLAVPI
     301  GVKFERFSDC KRGSYELTLA YVPDVIRKDP KSTATLASGA TWSTHGNNLS
     351  RQGLQLRLGN HCLINPGIEV FSHGAIELRG SSRNYNINLG GKRYRF*

```

This sequence is frame-shifted with respect to cp0018.

The cp0019 nucleotide sequence <SEQ ID 22> is:

```

15      1  TTGCAAGACT CTCAAGACTA TAGCTTTGTA AAGTTATCTC CAGGAGCGGG
      51  AGGGACTATA ATTACTCAAG ATGCTTCTCA GAAGCCTCTT GAAGTAGCTC
     101  CTTCTAGACC ACATTATGGC TATCAAGGAC ATTGGAATGT GCAAGTCATC
     151  CCAGGAACGG GAACTCAACC GAGCCAGGCA AATTTAGAAT GGGTGCGGAC
     20  201  AGGATACCTT CCGAATCCCG AACGGCAAGG ATCTTTAGTT CCCAATAGCC
     251  TGTGGGGTTC TTTTGTGAT CAGCGTGCTA TCCAAGAAAT CATGGTAAAT
     301  AGTAGCCAAA TCTTATGTCA GGAACGGGGA GTCTGGGGAG CTGGAATTGC
     351  TAATTTCCCTA CATAGAGATA AAATTAATGA GCACGGCTAT CGCCATAGCG
     401  GTGTCGGTTA TCTTGTGGGA GTTGGCACTC ATGCTTTTTC TGATGCTACG
     451  ATAAATGCGG CTTTTTGCCA GCTCTTCAGT AGAGATAAAG ACTACGTAGT
     50  501  ATCCAAAAAT CATGGAACCTA GCTACTCAGG GGTCTGATTT CTTGAGGATA
     551  CCCTAGAGTT TAGAAGTCCA CAGGGATTCT ATACTGATAG CTCCTCAGAA
     601  GCTTGCTGTA ACCAAGTCGT CACTATAGAT ATGCAGTTGT CTTACAGCCA
     651  TAGAAATAAT GATATGAAAA CCAAATACAC GACATATCCA GAAGCTCAGG
     701  GATCTTGGGC AAATGATGTT TTTGGTCTTG AGTTTGAGC GACTACATAC
     751  TACTACCTTA ACAGTACTTT TTTATTTGAT TACTACTCTC CGTTTCTCAG
     801  GCTGCAGTGC ACCTATGCTC ACCAGGAAGA CTTCAAAGAG ACAGGAGGTG
     851  AGGTTTCGTCA CTTTACTAGC GGAGATCTTT TCAATTTAGC AGTTCCTATT
     901  GGCGTGAAGT TTGAGAGATT TTCAGACTGT AAAAGGGGAT CTTATGAAC
     951  TACCCTTGCT TATGTTCTCG ATGTGATTCG CAAAGATCCC AAGAGCACGG
    100  1001  CAACATTGGC TAGTGGAGCT ACGTGGAGCA CCCACGGAAC CAATCTCTCC
     1051  AGACAAGGAT TACAAGTGGC TTTAGGGAAC CACTGTCTCA TAAATCCTGG
     1101  AATTGAGGTG TTCAGTCACG GAGCTATTGA ATTGCGGGGA TCCTCTCGTA
     1151  ATTATAACAT CAATCTCGGG GTAAATACC GATTTTAA

```

The PSORT algorithm predicts a cytoplasmic location (0.189).

40 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 11A. This protein was used to immunise mice, whose sera were used in a Western blot (Figure 11B) and an immunoblot assay (Figure 11C). A his-tagged protein was also expressed.

These experiments show that cp0019 is a useful immunogen. These properties are not evident from the sequence alone.

### 45 Example 12

The following *C.pneumoniae* protein (PID 4376466) was expressed <SEQ ID 23; cp6466>:

```

      1  MRKISVGICI TILLSLSVVL QGCKESSHSS TSGELAINI RDEPRSLDPR
     51  QVRLSEISL VKHIYEGLVQ ENNLSGNIEP ALAEDYSLSS DGLTYTFKLLK
    101  SAFWSNGDPL TAEDFIESWK QVATQEVSGI YAFALNPIKN VRKIQEGHLS
    151  IDHFGVHSPN ESTLVVTTLES PSHFLKLLA LPVFFPVHKS QRTLQSKSLP
    201  IASGAFYPKN IKQKQWIKLS KNPHYNNQSQ VETKTITIHF IPDANTAACL

```

251 FNQGKLNWQG PPWGERIPQE TLSNLQSKGH LHSFDVAGTS WLTFNINKFP  
 301 LNNMKLREAL ASALDKEALV STIFLGRAKT ADHLLPTNIH SYPEHQKQEM  
 351 AQRQAYAKKL FKEALEELQI TAKDLEHLNL IFPVSSSSASS LLVQLIREQW  
 401 KESLGFAPAI VGKEFALLQA DLSSGNFSLA TGGWFADFAD PMAFLTIFAY  
 451 PSGVPPYAIN HKDFLEILQN IEQEQDHQKR SELVSQASLY LETFHIIEPI  
 501 YHDAFQFAMN KKLSNLGVSP TGVVDFRYAK EN\*

A predicted signal peptide is highlighted.

The cp6466 nucleotide sequence <SEQ ID 24> is:

10 1 ATGCGCAAGA TATCAGTGGG AATCTGTATC ACCATTCTCC TTAGCCTCTC  
 51 CGTAGTCCTC CAAGGCTGCA AGGAGTCCAG TCACTCCTCT ACATCTCGGG  
 101 GAGAACTCGC TATTAATATA AGAGATGAAC CCCGTTCTTT AGATCCAAGA  
 151 CAAGTGCGAC TTCTTTCAGA AATCAGCCTT GTCAAACATA TCTATGAGGG  
 201 ATTAGTTCAA GAAAATAATC TTTCAGGAAA TATAGAGCCT GCTCTTGCAG  
 251 AAGACTACTC TCTTTCCTCG GACGGACTCA CTTATACTTT TAAACTGAAA  
 15 301 TCAGCTTTTT GGAGTAATGG CGACCCCTTA ACAGCTGAAG ACTTTATAGA  
 351 ATCTTGGAAA CAAGTAGCTA CTCAAGAAGT CTCAGGAATC TATGCTTTTG  
 401 CCTTGAATCC AATTAAAAAT GTACGAAAGA TCCAAGAGGG ACACCTCTCC  
 451 ATAGACCATT TTGGAGTGCA CTCTCCTAAT GAATCTACAC TTGTTGTTAC  
 501 CCTGGAATCC CCAACCTCGC ATTCTTTAAA ACTTTTAGCT CTTCCAGTCT  
 20 551 TTTTCCCCGT TCATAAATCT CAAAGAACCC TGCAATCCAA ATCTCTACCT  
 601 ATAGCAAGCG GAGCTTTCTA TCCTAAAAAT ATCAAACAAA AACAATGGAT  
 651 AAAACTCTCA AAAAACCCCTC ACTACTATAA TCAAAGTCAG GTGGAAACTA  
 701 AAACGATTAC GATTCAC TTCCTCCGATG CAAACACAGC AGCAAAACTA  
 751 TTTAATCAGG GAAAAC TCAAA TGGCAAGGA CCTCCTTGGG GAGAACGCAT  
 25 801 TCCTCAAGAA ACCCTATCCA ATTTACAGTC TAAGGGGCAC TTACACTCTT  
 851 TTGATGTCGC AGGAACCTCA TGGCTCACCT TCAATATCAA TAAATTCCCC  
 901 CTCACAATA TGAAGCTTAG AGAAGCCTTA GCATCAGCCT TAGATAAGGA  
 951 AGCTCTTGTC TCAACTATAT TCTTAGGCCG TGCAAAAACCT GCCGATCATC  
 1001 TCCTACCTAC AAATATTAT AGCTATCCCG AACATCAAAA ACAAGAGATG  
 30 1051 GCACAACGCC AAGCTTACGC TAAAAAACTC TTTAAAGAAG CTTTGAAGA  
 1101 ACTCCAAATC ACTGCTAAAG ATCTCGAACA TCTTAATCTT ATCTTTCCCG  
 1151 TTTCCCTCGTC AGCAAGTTCT TTACTAGTCC AACTTATACG AGAACAGTGG  
 1201 AAAGAAAGTT TAGGGTTCGC TATCCCTATT GTCGGAAAGG AATTTGCTCT  
 1251 TCTCCAAGCA GACCTATCTT CAGGGAACCT CTCTTTAGCT ACAGGAGGAT  
 35 1301 GGTTCGCAGA CTTTGCTGAT CCTATGGCAT TTCTAACGAT CTTTGCTTAT  
 1351 CCATCAGGAG TTCCTCCTTA TGCAATCAAC CATAAGGACT TCCTAGAAAT  
 1401 TCTACAAAAC ATAGAACAAG AGCAAGATCA CCAAAAACGC TCGGAATTAG  
 1451 TGTCGCAAGC TTCTCTTTAC CTAGAGACCT TTCATATTAT TGAGCCGATC  
 1501 TACCACGACG CATTTCAATT TGCTATGAAT AAAAAACTTT CTAATCTAGG  
 40 1551 AGTCTACCA ACAGGAGTTG TGGACTTCCG TTAGCTAAG GAAATTAG

The PSORT algorithm predicts that the protein is an outer membrane lipoprotein (0.790).

The protein was expressed in *E.coli* and purified both as a GST-fusion product and a His-tag fusion product. Purification of the protein as a GST-fusion product is shown in Figure 12A. The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 12B and 12C). FACS analysis was also performed.

These experiments show that cp6466 is a useful immunogen. These properties are not evident from the sequence alone.

### Example 13

The following *C.pneumoniae* protein (PID 4376468) was expressed <SEQ ID 25; cp6468>:

50 1 MFSRWITLFL LFISLTGCS YSSKHQSLI IPIHDDPVAF SPEQAKRAMD  
 51 LSIAQLLFDG LTRETHRESN DLELAIASRY TVSEDFCSYT FFIKDSALWS  
 101 DGTPIITSEDI RNAWEYAQEN SPHIQIFQGL NFSTPSSNAI TIHLDSPNPD  
 151 FPKLLAFPAF AIFKPENPKL FSGPYTLVEY FPGHNIHLK NPNNYDYHCV  
 201 SINSIKLLII PDIYTAIHL NRGKVDWVGQ PWHQGI PWEL HKQSQYHYT  
 55 251 YPVEGAFWLC LNTKSPHLND LQNRHRLATC IDKRSIIEEA LQGTQQPAET

301 LSRGAPQPNQ YKKQKPLTPQ EKLVLTYPSD ILRCQRIAEI LKEQWKAAGI  
 351 DLILEGLEHYH LFNVRKRVQD YAIATQTGVA YYPGANLISE EDKLLQNFEI  
 401 IPIYYLSYDY LTQDFIEGVI YNASGAVDLK YTYFP\*

A predicted signal peptide is highlighted.

5 The cp6468 nucleotide sequence <SEQ ID 26> is:

1 ATGTTTTTAC GATGGATCAC CCTCTTTTTA TTATTCATTA GCCTTACTGG  
 51 ATGCTCCTCC TACTCTTCAA AACATAAACA ATCTTTAATT ATTCCCATAC  
 101 ATGACGACCC TGTAGCTTTT TCTCCTGAAC AAGCAAAACG GGCCATGGAC  
 151 CTTTCTATTG CCCAACTTCT TTTTGATGGT CTGACTAGAG AAACATATCG  
 201 CGAATCCAAT GATTTTGAAT TAGCGATTGC CAGTCGCTAT ACAGTCTCTG  
 251 AAGACTTTTG CTCTTATACG TTCTTTATCA AAGACAGCGC TTTATGGAGC  
 301 GACGGAACAC CAATCACCTC CGAAGATATC CGTAACGCTT GGGAGTATGC  
 351 ACAGGAGAAC TCTCCCACA TACAGATCTT CCAAGGACTT AACTTCTCAA  
 401 CTCCTTCATC AAATGCAATT ACGATTATC TCGACTCGCC CAACCCCGAT  
 451 TTTCTTAAGC TTCTTGCTTT TCCTGCATTT GCTATCTTTA AACCAGAAAA  
 501 CCCGAAGCTC TTTAGCGGTC CGTATACTCT TGTAGAGTAT TTCCCAGGGC  
 551 ATAACATTC AATTAAAGAAA AACCTAACT ATTACGACTA CCACTGCGTC  
 601 TCCATCAACT CCATCAAAC GCTCATTTAT CCTGATATAT ATACAGCCAT  
 651 CCACCTCCTA AACAGAGGCA AGGTGGACTG GGTAGGACAA CCCTGGCATC  
 701 AAGGGATTCC TTGGGAGCTC CATAAACAAT CGCAATATCA CTACTACACC  
 751 TATCCTGTAG AAGGTGCTTT CTGGCTTTGT CTAAATACAA AATCCCCACA  
 801 CTTAAATGAT CTTCAAAACA GACATAGACT CGTACTTGT ATTGATAAAC  
 851 GTTCTATCAT TGAAGAAGCT CTTCAAGGAA CCAACAACC AGCGGAAACA  
 901 CTGTCCCGAG GAGCTCCACA ACCAAATCAA TATAAAAAAC AAAAGCCTCT  
 951 AACTCCACAA GAAAACTCG TGCTTACCTA TCCCTCAGAT ATTCTAAGAT  
 1001 GCCAACGCAT AGCAGAAATC TTAAAGGAAC AATGGAAAGC TGCTGGAATA  
 1051 GATTTAATCC TTGAAGGACT CGAATACCAT CTGTTTGTTA ACAAACGAAA  
 1101 AGTCCAAGAC TACGCCATAG CAACACAGAC TGGAGTTGCT TATTACCCAG  
 1151 GAGCAAATCT AATTTCTGAA GAAGACAAGC TCCTGCAAAA CTTTGAGATT  
 1201 ATCCCGATCT ACTATCTGAG CTATGACTAT CTCACTCAAG ATTTTATAGA  
 1251 GGGAGTAATC TATAATGCTT CTGGAGCTGT AGATCTCAAA TATACCTATT  
 1301 TCCCCTAG

The PSORT algorithm predicts that this protein is an outer membrane lipoprotein (0.790).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 13A.

35 The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 13B) and for FACS analysis. A his-tagged protein was also expressed.

These experiments show that cp6468 is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 14

40 The following *C.pneumoniae* protein (PID 4376469) was expressed <SEQ ID 27; cp6469>:

1 MKMHLKPTL KSLIPNLLFL LLTLSSCSKQ KQEPLGKHLV IAMSHDLADL  
 51 DPRNAYLSRD ASLAKALYEG LTRETDQGIA LALAESYTLS KDHKVYTFKL  
 101 RPSVWSDGTP LTAYDFEKSI KQLYFEEFSP SIHTLLGVIK NSSAIHNAOK  
 151 SLETLGIAQAK DDLTLVITLE QPFPYFLTLI ARPVFSPVHH TLRESYKKG  
 201 PPSTYISNGP FVLKKHEHQY YLILEKNPHY YDHESVKLDR VTLKIIPDAS  
 251 TATKLFKSKS IDWIGSPWSA PISNEDQKVL SQEKILTYSV SSTTLLIYNL  
 301 QKPLIQNKAL RKAIAHAIDR KSILRLVPSG QEAVTLVPPN LSQNLQKEI  
 351 STEERQTKAR AYFQEAQETL SEKELAEISI LYPIDSSNSS IIAQEIQRQL  
 401 KDTLGLKIKI QGMEYHCFK KRRQGDFFIA TGGWIAEYVS PVAFLSILGN  
 451 PRDLTQWRNS DYEKHTLEKLY LPHAYKENLK RAEMIIEEET PIIPLYHGKY  
 501 IYAIHPKIQN TFGSLLGHTD LKNIDILS\*

A predicted signal peptide is highlighted.

The cp6469 nucleotide sequence <SEQ ID 28> is:

```

      1 ATGAAGATGC ATAGGCTTAA ACCTACCTTA AAAAGTCTGA TCCCTAATCT
    51 TCTTTTCTTA TTGCTCACTC TTTCAAGCTG CTCAAAGCAA AAACAAGAAC
   101 CCTTAGGAAA ACATCTCGTT ATTGCGATGA GCCATGATCT CGCCGACCTA
   151 GATCCTCGCA ATGCCATATT AAGCAGAGAT GCTTCCCTAG CAAAAGCCCT
5    201 CTATGAAGGA CTGACAAGAG AAACCTGATCA AGGAATCGCA CTGGCTCTTG
   251 CAGAAAGTTA TACCTGTGCA AAAGATCATA AGGTCTATAC CTTTAAACTC
   301 AGACCTTCTG TGTGGAGCGA TGGCACTCCA CTCACTGCTT ATGACTTTGA
   351 AAAATCTATA AAACAACTGT ACTTCGAAGA ATTTTCACCT TCCATACATA
   401 CTTTACTCGG CGTGATTAAA AATCTTTCGG CAATCCACAA TGCTCAAAAA
  10   451 TCTCTGGAAA CTCTTGGGAT ACAGGCAAAA GATGATCTTA CTTTGGTGAT
   501 TACCTTAGAG CAACCTTTCC CATACTTTCT CACACTTATC GCTCGCCCCG
   551 TATTCTCCCC TGTTTCATCAC ACCCTTAGGG AATCCTATAA GAAAGGAACA
   601 CCCCCATCCA CATACTCTC CAATGGGCCC TTTGTCTTAA AAAACATGA
   651 ACACCAAAAC TACTTAATTT TAGAAAAAAA TCCTCACTAC TATGATCATG
  15   701 AATCAGTAAA GTTAGACCGA GTCACCTTAA AAATTATCCC AGACGCCCTC
   751 ACAGCCACGA AACTTTTCAA AAGTAAATCT ATAGATTGGA TTGGCTCACC
   801 TTGGAGCGCT CCGATATCTA ACGAAGACCA AAAAGTTCTC TCCCAAGAAA
   851 AGATCTTAC CTATTCTGTT TCAAGCACCA CCCTTCTTAT CTATAACCTG
   901 CAAAAACCTC TAATACAAAA TAAAGCCCTC AGGAAAGCCA TTGCTCATGC
  20   951 TATTGATAGA AAATCTATCT TAAGACTCGT GCCTTCAGGA CAAGAAGCTG
  1001 TAACTCTAGT TCCCCCAAAT CTTTCACAAC TCAATCTTCA AAAAGAGATC
  1051 TCAACAGAAG AACGACAAAC AAAAGCCAGA GCATATTTTC AAGAAGCTAA
  1101 AGAAACACTT TCTGAAAAAG AACTCGCAGA ACTCAGCATC CTCATCCTA
  1151 TAGATTCTCT GAATTCCTCC ATCATAGCTC AAGAAATCCA AAGACAACCT
  25   1201 AAGATACCT TAGGATTGAA AATCAAAATC CAAGGCATGG AGTACCCTG
  1251 CTTTTTAAAG AAACGTCGTC AAGGAGATTT CTTTCATAGCG ACAGGAGGAT
  1301 GGATTGCGGA ATACGTAAGC CCCGTAGCCT TCCTATCTAT TCTAGGCAAC
  1351 CCCAGAGACC TCACACAATG GAGAAACAGT GATTACGAAA AGACTTTAGA
  1401 GAAACTCTAT CTCCCTCATG CCTACAAAGA GAATTTAAAA CGCGCAGAAA
  30   1451 TGATAATAGA AGAAGAAACC CCGATTATCC CCCTGTATCA CGGCAAAATAT
  1501 ATTTACGCTA TACATCTTAA AATCCAGAA ACATTGCGAT CTCTTCTAGG
  1551 CCACACAGAT CTCAAAAATA TCGATATCTT AAGTTAG

```

The PSORT algorithm predicts a periplasmic location (0.934).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 14A.

35 The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 14B) and for FACS analysis. A his-tagged protein was also expressed.

These experiments show that cp6469 is a useful immunogen. These properties are not evident from the sequence alone.

### Example 15

40 The following *C.pneumoniae* protein (PID 4376602) was expressed <SEQ ID 29; cp6602>:

```

      1 MAASGGTGGL GGTQGVNLAA VEAAAAKADA AEVVASQEGS EMNMIQQSQD
      51 LTNPAAATRT KKKEEFQTL ESRKKGEAGK AEKKSESTEE KPDTDLADKY
     101 ASGNSEISGQ ELRGLRDAIG DDASPEDILA LVQEKIKDPA LQSTALDYLV
     151 QTTPPSQGKL KEALIQARNT HTEQFGRTAI GAKNILFASQ EYADQLNVSP
45    201 SGLRSLYLEV TGDTHTCQDL LSMLQDRYTY QDMAIVSSFL MKGMATELKR
     251 QGPYVPSAQL QVLMETERNL QAVLTSYDYF ESRVPILLDS LKAEGIQTPS
     301 DLNFKVVAES YHKIINDKFP TASKVEREVR NLIGDDVDSV TGVNLNLFSA
     351 LRQTSRLFS SADKRQQLGA MIANALDAVN INNEDYPKAS DFPKPYPWS*

```

The cp6602 nucleotide sequence <SEQ ID 30> is:

```

    50      1 ATGGCAGCAT CAGGAGGCAC AGGTGGTTTA GGAGGCACTC AGGGTGTCAA
      51 CCTTGCAGCT GTAGAAGCTG CAGCTGCAAA AGCAGATGCA GCAGAAGTTG
     101 TAGCCAGCCA AGAAGGTTCT GAGATGAACA TGATTCAACA ATCTCAGGAC
     151 CTGACAAATC CCGCAGCAGC AACACGCACG AAAAAAAGG AAGAGAAGTT
     201 TCAAACCTTA GAATCTCGGA AAAAAGGAGA AGCTGGAAG GCTGAGAAAA
55    251 AATCTGAATC TACAGAAGAG AAGCCTGACA CAGATCTTGC TGATAAGTAT
     301 GCTTCTGGGA ATTCTGAAAT CTCTGGTCAA GAACTTCGCG GCCTGCGTGA
     351 TGCAATAGGA GACGATGCTT CTCCAGAAGA CATTCTTGCT CTTGTACAG

```

```

401 AGAAAATTAA AGACCCAGCT CTGCAATCCA CAGCTTTGGA CTACCTGGTT
451 CAAACGACTC CACCCTCCCA AGGTAAATTA AAAGAAGCGC TTATCCAAGC
501 AAGGAATACT CATACGGAGC AATTCGGACG AACTGCTATT GGTGCGAAAA
551 ACATCTTATT TGCCTCTCAA GAATATGCAG ACCAACTGAA TGTTCCTCCT
601 TCAGGGCTTC GCTCTTTGTA CTTAGAAGTG ACTGGAGACA CACATACCTG
651 TGATCAGCTA CTTTCTATGC TTCAAGACCG CTATACCTAC CAAGATATGG
701 CTATTGTCAG CTCCTTTCTA ATGAAAGGAA TGGCAACAGA ATTA AAAAGG
751 CAGGGTCCCT ACGTACCCAG TCGCAACTA CAAGTTCTCA TGACAGAAAC
801 TCGTAACCTG CAAGCAGTTC TTACCTCGTA CGATTACTTT GAAAGTCGCG
851 TTCTATTTT ACTCGATAGC TTA AAAAGCTG AGGGAATCCA AACTCCTTCT
901 GATCTAAACT TTGTGAAGGT AGCTGAGTCC TACCATAAAA TCATTAACGA
951 TAAGTTCCCA ACAGCATCTA AAGTAGAACG AGAAGTCCGC AATCTCATAG
1001 GAGACGATGT TGATTCTGTG ACCGGTGTCT TGAACCTATT CTTTTCTGCT
1051 TTACGTCAA CCGTCGTCAG CCTTTTCTCT TCAGCAGACA AACGTCAGCA
1101 ATTAGGAGCT ATGATTGCTA ATGCTTTAGA TGCTGTAAAT ATAAACAATG
1151 AAGATTATCC CAAAGCATCA GACTTCCCTA AACCTATCC TTGGTCATGA

```

The PSORT algorithm predicts a cytoplasmic location (0.080).

The protein was expressed in *E.coli* and purified as both a His-tag and a GST-fusion product, as shown in Figure 15A. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 15B) and for FACS analysis (Figure 15C).

The cp6602 protein was also identified in the 2D-PAGE experiment (Cpn0324).

These experiments show that cp6602 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 16

The following *C.pneumoniae* protein (PID 4376727) was expressed <SEQ ID 31; cp6727>:

```

1  MKYSLPWLLT SSALVFSLHP LMAANTDLSS SDNYENGSSG SAAFTAKETS
51 DASGTTYTLT SDVSITNVSA ITPADKSCFT NTGGALSFVG ADHSLVLQTI
101 ALTHDGAAIN NTNTALSFSG FSSLLIDSAP ATGTSGGKGA ICVTNTEGGT
151 ATFTDNASVT LQKNTSEKDG AAVSAYSIDL AKTTTAALLD QNTSTKNNGA
201 LCSTANTTVQ GNSGTVTFFS NTATDKGGGI YSKEKDSTLD ANTGVVTFKS
251 NTAKTGGAWS SDDNLALTGN TQVLFQENKT TGSAAQANNP ECGCGAICCY
301 LATATDKTGL AISQNQEMSF TSNTTTANGG AIYATKCTLD GNTTLTFDQN
351 TATAGCGGAI YTETEDFSK GSTGTVTFFS NTAKTGGALY SKGNSSLTGN
401 TNLLFSGNKA TGPSNSSANQ ECGCGAILAF IDSGSVSDKT GLSIANNQEV
451 SLTSNAATVS GGAIYATKCT LTGNGLTFD GNTAGTSGGA IYTETEDFTL
501 TGSTGTVTFS TNAKTGGAL YSKGNNSLSG NTNLLFSGNK ATGPSNSSAN
551 QEGCGGAILS FLESASVSTK KGLWIEDNEN VSLSGNTATV SGGAIYATKC
601 ALHGNTTLTF DGNTAETAGG AIYTETEDFT LTGSTGTVTF STNTAKTAGA
651 LHTKGNTSFT KNKALVFSGN SATATATTTT DQEGCGGAIL CNISESDIAT
701 KSLTLTENES LSFINNTAKR SGGGIYAPKC VISGSESINF DGNTAETSGG
751 AIYSKNLSIT ANG PVSFTNN SGGKGGAIYI ADSELSLEA IDGDITFSGN
801 RATEGTSTPN SIHLGAGAKI TKLAAAPGHT IYFYDPITME APASGGTIEE
851 LVINPVVKAI VPPQPKNP IASVPVVPVA PANPNTGTIV FSSGKLPSQD
901 ASIPANTTTI LNQKINLAGG NVVLKEGATL QVVSFTQQPD STVFMDAGTT
951 LETTTTNTND GSIDLKNSLV NLDALDGKRM ITIAVNSTSG GLKISGDLKF
1001 HNEGSEFYDN PGLKANLNL PFLDLSSTSGT VNLDDEFNPIP SSMAAPDYGY
1051 QGSWTLVPKV GAGGKVTLVA EWQALGYTPK PELRATLVPN SLWNAYVINI
1101 SIQQEIAATAM SDAPSHPGIW IGGIGNAFHQ DKQKENAGFR LISRGYIVGG
1151 SMTTPQEYTF AVAFSQLFGK SKDYVVS DIK SQVYAGSLCA QSSYVIPLHS
1201 SLRRHVLSKV LPELPGETPL VLHGQVSYGR NHHNMTTKLA NNTQKSDWD
1251 SHSFAVEVGG SLPVDLNYRY LTSYSPYVKL QVSVNQKGF QEVAADPRIF
1301 DASHLVN VSI PMGLTFKHES AKPPSALLLT LGYAVDAYRD HPHCLTSLTN
1351 GTSWSTFATN LSRQAFFAEA SGHLKLLHGL DCFASGSCEL RSSRSRYNAN
1401 CGTRYSF*

```

A predicted signal peptide is highlighted.

The cp6727 nucleotide sequence <SEQ ID 32> is:

```

1  ATGAAATATT CTTTACCTTG GCTACTTACC TCTTCGGCTT TAGTTTTCTC
51  CCTACATCCA CTAATGGCTG CTAACACGGA TCTCTCATCA TCCGATAACT
101  ATGAAATGG TAGTAGTGGT AGCGCAGCAT TCACTGCCAA GGAAACTTCG
5   151  GATGCTTCAG GAACTACCTA CACTCTCACT AGCGATGTTT CTATTACGAA
201  TGTATCTGCA ATTACTCCTG CAGATAAAAG CTGTTTACAA AACACAGGAG
251  GAGCATTGAG TTTTGTGTGA GCTGATCACT CATTTGGTCT GCAAACCATA
301  GCGCTTACGC ATGATGGTGC TGCAATTAAC AATACCAACA CAGCTCTTTC
10  351  TTTCTCAGGA TTCTCGTCAC TCTTAATCGA CTCAGCTCCA GCAACAGGAA
401  CTTCCGGCGG CAAGGTGCT ATTTGTGTGA CAAATACAGA GGGAGGTACT
451  GCGACTTTTA CTGACAAATG CAGTGTCACT CTCCAAAAAA ATACTTCAGA
501  AAAAGATGGA GCTGCAGTTT CTGCCCTACAG CATCGATCTT GCTAAGACTA
551  CGACAGCAGC TCTCTTAGAT CAAAATACTA GCACAAAAAA TGGCGGGGCC
601  CTCTGTAGTA CAGCAAACAC TACAGTCCAA GGAAACTCAG GAACGGTGAC
15  651  CTTCTCCTCA AATACTGTCA CAGATAAAGG TGGGGGGATC TACTCAAAG
701  AAAAGGATAG CACGCTAGAT GCCAATACAG GAGTCGTTAC CTTCAAATCT
751  AATACTGCAA AGACGGGGGG TGCTTGGAGC TCTGATGACA ATCTTGCTCT
801  TACCGGCAAC ACTCAAGTAC TTTTTCAGGA AAATAAAACA ACCGGCTCAG
851  CAGCACAGGC AAATAACCCG GAAGGTTGTG GTGGGGCAAT CTGTTGTTAT
20  901  CTTGCTACAG CAACAGACAA AACTGGATTA GCCATTTCTC AGAATCAAGA
951  AATGAGCTTC ACTAGTAATA CAACAACCTG GAATGGTGA GCGATCTACG
1001  CTACTAAATG TACTCTGGAT GGAAACACAA CTCTTACCTT CGATCAGAAT
1051  ACTGCGACAG CAGGATGTGG CGGAGCTATC TATACAGAAA CTGAAGATTT
1101  TTCTCTTAAG GGAAGTACGG GAACCGTGAC CTTACAGACA AATACAGCAA
25  1151  AGACAGGCGG CGCCTTATAT TCTAAAGGAA ACAGCTCGCT GACTGGAAAT
1201  ACCAACCTGC TCTTTTCAGG GAACAAAGCT ACGGGCCCGA GTAATTTCTT
1251  AGCAAATCAA GAGGGTTGCG GTGGGGCAAT CCTAGCCTTT ATTGATTCAG
1301  GATCCGTAAG CGATAAAACA GGAATATCGA TTGCAAACAA CCAAGAAGTC
1351  AGCCTCACTA GTAATGCTGC AACAGTAAGT GGTGGTGCGA TCTATGCTAC
30  1401  CAAATGTACT CTAAGTGGAA ACGGCTCCCT GACCTTTGAC GGCAATACTG
1451  CTGGAACCTC AGGAGGGGCG ATCTATACAG AAAGTGAAGA TTTTACTCTT
1501  ACAGGAAGTA CAGGAACCGT GACCTTCAGC ACAAATACAG CAAAGACAGG
1551  CGCGCCTTAA TATTCTAAAG GCAACAACCT TCTGTCTGGT AATACCAACC
1601  TGCTCTTTTC AGGGAACAAA GCTACGGGCC CGAGTAATTC TTCAGCAAAAT
35  1651  CAAGAGGGTT GCGGTGGGGC AATCCTATCG TTTCTTGAGT CAGCATCTGT
1701  AAGTACTAAA AAAGGACTCT GGATTGAAGA TAACGAAAC GTGAGTCTCT
1751  CTGGTAATAC TGCAACAGTA AGTGGCGGTG CGATCTATGC GACCAAGTGT
1801  GCTCTGCATG GAAACACGAC TCTTACCTTT GATGGCAATA CTGCCGAAAC
1851  TGCAGGAGGA GCGATCTATA CAGAAACCGA AGATTTTACT CTTACGGGAA
40  1901  GTACGGGAAC CGTGACCTTC AGCACAATA CAGCAAAGAC AGCAGGGGCT
1951  CTACATACTA AAGGAAATAC TTCCTTTACC AAAAATAAGG CTCTTGATTT
2001  TTCTGGAAAT TCAGCAACAG CAACAGCAAC AACAACCTACA GATCAAGAAG
2051  GTTGTGGTGG AGCGATCCTC TGTAATATCT CAGAGTCTGA CATAGCTACA
2101  AAAAGCTTAA CTCTTACTGA AAATGAGAGT TTAAGTTTCA TTAACAATAC
45  2151  GGCAAAAAGA AGTGGTGGTG GTATTTATGC TCCTAAGTGT GTAATCTCAG
2201  GCAGTGAATC CATAAACTTT GATGGCAATA CTGCTGAAAC TTCGGGAGGA
2251  GCGATTTATT CGAAAAACCT TTCGATTACA GCTAACGGTC CTGTCTCCTT
2301  TACCAATAAT TCTGGAGGCA AGGGAGGCGC CATTTATATA GCCGATAGCG
2351  GAGAACTTTC CTTAGAGGCT ATTGATGGGG ATATTACTTT CTCAGGGAAC
50  2401  CGAGCGACTG AGGGAACCTC AACTCCCAAC TCGATCCATT TAGGTGCAGG
2451  GGCTAAGATC ACTAAGCTTG CAGCAGCTCC TGGTCATACG ATTTATTTT
2501  ATGATCCTAT TACGATGGAA GCTCCTGCAT CTGGAGGAAC AATAGAGGAG
2551  TTAGTCATCA ATCCTGTTGT CAAAGCTATT GTTCTCTCTC CCCAACCAA
2601  AAATGGTCCT ATAGCTTCAG TGCTGTAGT CCCTGTAGCA CCTGCAACC
55  2651  CAAACACGGG AACTATAGTA TTTTCTTCTG GAAACTCCC CAGTCAAGAT
2701  GCCTCGATTC CTGCAAATAC TACCACCATA CTGAACCAGA AGATCAACTT
2751  AGCAGGAGGA AATGTCGTTT TAAAAGAAGG AGCCACCCTA CAAGTATATT
2801  CCTTCACACA GCAGCCTGAT TCTACAGTAT TCATGGATGC AGGAACGACC
2851  TTAGAGACCA CGACAACCTA CAATACAGAT GGCAGCATCG ATCTAAAGAA
60  2901  TCTCTCTGTA AATCTGGATG CTTTAGATGG CAAGCGTATG ATAACGATTG
2951  CCGTAAACAG CACAAGTGGG GGATTAAAAA TCTCAGGGGA TCTGAAATTC
3001  CATAACAATG AAGGAAGTTT CTATGACAAT CCTGGGTTGA AAGCAAACCT
3051  AAATCTTCCT TTCTTAGATC TTTCTTCTAC TTCAGGAAC TAAATTTTAG
65  3101  ACGACTTCAA TCCGATTCC TCTAGCATGG CTGCTCCGGA TTATGGGTAT
3151  CAAGGGAGTT GCACTCTGGT TCCTAAAGTA GGAGCTGGAG GGAAGGTGAC
3201  TTTGGTCGCG GAATGGCAAG CGTTAGGATA CACTCCTAAA CCAGAGCTTC
3251  GTGCGACTTT AGTTCCTAAT AGCCTTTGGA ATGCTTATGT AAACATCCAT

```

```

3301 TCTATACAGC AGGAGATCGC CACTGCGATG TCGGACGCTC CCTCACATCC
3351 AGGGATTTGG ATTGGAGGTA TTGGCAACGC CTTCCATCAA GACAAGCAAA
3401 AGGAAAATGC AGGATTCGGT TTGATTTCCA GAGGTTATAT TGTTGGTGGC
5 3451 AGCATGACCA CCCCTCAAGA ATATACCTTT GCTGTTGCAT TCAGCCAACT
3501 CTTTGGCAAA TCTAAGGATT ACGTAGTCTC GGATATTAAA TCTCAAGTCT
3551 ATGCAGGATC TCTCTGTGCT CAGAGCTCTT ATGTCATTCC CCTGCATAGC
3601 TCATTACGTC GCCACGTCCT CTCTAAGGTC CTTCCAGAGC TCCCAGGAGA
3651 AACTCCCCTT GTTCTCCATG GTCAAGTTTC CTATGGAAGA AACCACCATA
10 3701 ATATGACGAC AAAGCTTGCG AACAAACAC AAGGGAAATC AGACTGGGAC
3751 AGCCATAGCT TCGCTGTGTA AGTCGGTGGT TCTCTTCCTG TAGATCTAAA
3801 CTACAGATAC CTTACCAGCT ACTCTCCCTA TGTGAAACTC CAAGTTGTGA
3851 GTGTAAATCA AAAAGGATT CAGAGGTTG CTGCTGATCC ACGTATCTTT
3901 GACGCTAGCC ATCTGGTCAA CGTGTCTATC CCTATGGGAC TCACCTTCAA
15 3951 ACACGAATCA GCAAAGCCCC CCAGTGCTTT GCTTCTTACT TTAGGTTACG
4001 CTGTAGATGC TTACCGGGAT CACCTCACT GCCTGACCTC CTTAACAAAT
4051 GGCACCTCGT GGTCTACGTT TGCTACAAAC TTATCACGAC AAGCTTCTTT
4101 TGCTGAGGCT TCTGGACATC TGAAGTACT TCATGGTCTT GACTGCTTCG
4151 CTTCTGGAAG TTGTGAACTG CGCAGCTCCT CAAGAAGCTA TAATGCAAAAC
4201 TGTGGAATC GTTATTCTTT CTAA

```

20 The PSORT algorithm predicts an outer membrane location (0.915).

The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 16A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 16B) and for FACS analysis (Figure 16C). A GST-fusion protein was also expressed.

The cp6727 protein was also identified in the 2D-PAGE experiment (Cpn0444).

25 These experiments show that cp6727 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 17

The following *C.pneumoniae* protein (PID 4376731) was expressed <SEQ ID 33; cp6731>:

```

30 1 MKSSLHWFLI SSSLALPLSL NFSFAA VVE INLGPTNSFS GPGTYTPPAQ
51 TTNADGTIYN LTGDVSI TNA GSPTALTASC FKETTGNLSF QGHGYQFLLQ
101 NIDAGANCTF TNTAANKLLS FSGFSYLSLI QTTNATTGTG AIKSTGACSI
151 QSNYSCYFGQ NFSNDNGGAL QGSSISLSLN PNLTFKNKA TQKGGALYST
201 GGITINNTLN SASFSENTAA NNGGAIYTEA SSFISNKAI SFINNSVTAT
35 251 SATGGAIYCS STSAPKPVLT LSDNGELNFI GNTAITSGGA IYTDNLVLSS
301 GGPTLTKNNS AIDTAAPLGG AIAIADSGSL SLSALGGDIT FEGNTVVKGA
351 SSSQTTTRNS INIGNTNAKI VQLRASQNT IYFYDPITTS ITAALSDALN
401 LNGPDLAGNP AYQGTIVFSG EKLSEAEAAE ADNLKSTIQQ PLTLAGQQLS
451 LKSGVTLVAK SFSQSPGSTL LMDAGTTLET ADGITINNLV LNVDLSKETK
501 KATLKATQAS QTVTLSGSL S LVDPSGNVYE DVSWNNPQVF SCLTLTADDP
40 551 ANIHITDLAA DPLEKNPIHW GYQGNWALSW QEDTATKSKA ATLTWTKTGY
601 NPNPERRGTL VANTLWGSFV DVRSIQQLVA TKVRQSQETR GIWCEGISNF
651 FHKDSTKINK GFRHISAGYV VGATTTLASD NLITAFCQL FGKDRDHFIN
701 KNRASAYAAS LHLQLATLS SPSLLRLYPG SESEQPVLEF AQISYIYSKN
751 TMKTYTQAP KGESSWYNDG CALELASSLP HTALSHEGLF HAYFPFIKVE
45 801 ASYIHQDSFK ERNTTLVRSF DSGDLINVS PIGITFERFS RNERASYEAT
851 VIYVADVYRK NPDCTTALLI NNTSWKTTGT NLSRQAGIGR AGIFYAFSPN
901 LEVTSNLSME IRGSSRSYNA DLGGKQF*

```

A predicted signal peptide is highlighted.

The cp6731 nucleotide sequence <SEQ ID 34> is:

```

50 1 ATGAAATCCT CTCTTCATTG GTTTTTAATC TCGTCATCTT TAGCACTTCC
51 CTTGTCACTA AATTTCTCTG CGTTTGCTGC TGTGTTGAA ATCAATCTAG
101 GACCTACCAA TAGCTTCTCT GGACCAGGAA CCTACACTCC TCCAGCCCAA
151 ACAACAAATG CAGATGGAAC TATCTATAAT CTAACAGGGG ATGCTCTCAAT
201 CACCAATGCA GGATCTCCGA CAGCTCTAAC CGCTTCCTGC TTAAAGAAA

```

```

251 CTACTGGGAA TCTTTCTTTC CAAGGCCACG GCTACCAATT TCTCCTACAA
301 AATATCGATG CGGGAGCGAA CTGTACCTTT ACCAATACAG CTGCAAATAA
351 GCTTCTCTCC TTTTCAGGAT TCTCCTATTT GTCACTAATA CAAACCACGA
401 ATGCTACCAC AGGAACAGGA GCCATCAAGT CCACAGGAGC TTGTTCTATT
5   451 CAGTCGAACT ATAGTTGCTA CTTTGGCCAA AACTTTTCTA ATGACAATGG
501 AGGCGCCCTC CAAGGCAGCT CTATCAGTCT ATCGCTAAAC CCCAACCTAA
551 CGTTTGCCAA AAACAAAGCA ACGCAAAAAG GGGGTGCCCT CTATTCACAG
601 GGAGGGATTA CAATTAACAA TACGTTAAAC TCAGCATCAT TTTCTGAAAA
651 TACCGCGGCG AACAAATGGCG GAGCCATTTA CACGGAAGCT AGCAGTTTAA
10  701 TTAGCAGCAA CAAAGCAATT AGCTTTATAA ACAATAGTGT GACCGCAACC
751 TCAGCTACAG GGGGAGCCAT TTACTGTAGT AGTACATCAG CCCCCAAACC
801 AGTCTTAACT CTATCAGACA ACGGGGAAC TGAACCTTATA GGAAATACAG
851 CAATTACTAG TGGTGGGGCG ATTTATACTG ACAATCTAGT TCTTTCTTCT
901 GGAGGACCTA CGCTTTTATA AAACAACCTCT GCTATAGATA CTGCAGCTCC
15  951 CTTAGGAGGA GCAATTGCGA TTGCTGACTC TGGATCTTTG AGTCTTTCGG
1001 CTCTTGGTGG AGACATCACT TTTGAAGGAA ACACAGTAGT CAAAGGAGCT
1051 TTTTCGAGTC AGACCACTAC CAGAAATCTT ATTAACATCG GAAACACCAA
1101 TGCTAAGATT GTACAGCTGC GAGCCTCTCA AGGCAATACT ATCTACTTCT
1151 ATGATCCTAT AACAAC TAGC ATCACTGCAG CTCTCTCAGA TGCTCTAAAC
20  1201 TTAAATGGTC CTGACCTTGC AGGGAATCCT GCATATCAAG GAACCATCGT
1251 ATTTTCTGGA GAGAAGCTCT CGGAAGCAGA AGCTGCAGAA GCTGATAATC
1301 TCAAATCTAC AATTCAGCAA CCTCTAATC TTGCGGGAGG GCAACTCTCT
1351 CTTAATCAG GAGTCACTCT AGTTGCTAAG TCCTTTTCGC AATCTCCGGG
25  1401 CTCTACCCCTC CTCATGGATG CAGGGACCAC ATTAGAAACC GCTGATGGGA
1451 TCACTATCAA TAATCTTGTT CTCAATGTAG ATTCCTTAAA AGAGACCAAG
1501 AAGGCTACGC TAAAAGCAAC ACAAGCAAGT CAGACAGTCA CTTTATCTGG
1551 ATCGCTCTCT CTTGTAGATC CTTCTGGAAA TGTCTACGAA GATGTCTCTT
1601 GGAATAACCC TCAAGTCTTT TCTTGTCTCA CTCTTACTGC TGACGACCCC
30  1651 GCGAATATTC ACATACAGA CTTAGCTGCT GATCCCCTAG AAAAAATCC
1701 TATCCATTGG GGATACCAAG GGAATTGGGC ATTATCTTGG CAAGAGGATA
1751 CTGCGACTAA ATCCAAAGCA GCGACTCTTA CCTGGACAAA AACAGGATAC
1801 AATCCGAATC CTGAGCGTCG TGGAACCTTA GTTGCTAACA CGCTATGGGG
1851 ATCCTTTGTT GATGTGCGCT CCATACAACA GCTTGTAGCC ACTAAAGTAC
35  1901 GCCAATCTCA AGAACTCGC GGCATCTGGT GTGAAGGGAT CTCGAACCTC
1951 TTCCATAAAG ATAGCACGAA GATAAATAAA GGTTTTCGCC ACATAAGTGC
2001 AGGTTATGTT GTAGGAGCGA CTACAACATT AGCTTCTGAT AATCTTATCA
2051 CTGCAGCCTT CTGCCAATTA TTCGGGAAAAG ATAGAGATCA CTTTATAAAT
2101 AAAAATAGAG CTTCTGCCTA TGCAGCTTCT CTCCATCTCC AGCATCTAGC
40  2151 GACCTTGTCT TCTCCAAGCT TGTTACGCTA CCTTCCTGGA TCTGAAAGTG
2201 AGCAGCCTGT CCTCTTTGAT GCTCAGATCA GCTATATCTA TAGTAAAAAT
2251 ACTATGAAAA CCTATTACAC CCAAGCACCA AAGGGAGAGA GCTCGTGGTA
2301 TAATGACGGT TGCCTCTGG AACTTGCGAG CTCCTTACCA CACACTGCTT
2351 TAAGCCATGA GGGTCTCTTC CACGCGTATT TTCCTTTCAT CAAAGTAGAA
45  2401 GCTTCGTACA TACACCAAGA TAGCTTCAAA GAACGTAAATA CTACCTTGCT
2451 ACGATCTTTC GATAGCGGTG ATTTAATTAA CGTCTCTGTG CCTATTGGAA
2501 TTACCTTCGA GAGATTCTCG AGAAACGAGC GTGCGTCTTA CGAAGCTACT
2551 GTCATCTACG TTGCCGATGT CTATCGTAAG AATCCTGACT GCACGACAGC
2601 TCTCCTAATC AACAATACCT CGTGGAAC TACAGGAACG AATCTCTCAA
50  2651 GACAAGCTGG TATCGGAAGA GCAGGGATCT TTTATGCCTT CTCTCCAAAT
2701 CTTGAGGTCA CAAGTAACCT ATCTATGGAA ATTCTGGAT CTTACGCGAG
2751 CTACAATGCA GATCTTGGAG GTAAGTTCCA GTTCTAA

```

The PSORT algorithm predicts an outer membrane location (0.926).

The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 17A. A GST-fusion protein was also expressed. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 17B; his-tag) and for FACS analysis (Figure 17C; his-tag and GST-fusion).

The GST-fusion protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis. Less cross-reactivity was seen with the his-fusion.



These experiments show that cp6731 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 18

The following *C.pneumoniae* protein (PID 4376737) was expressed <SEQ ID 35; cp6737>:

```

5      1  MPLSFKSSSF CLLACLCSAS CAFAETRLGG NRVPPITNQG EEILLTSDFV
      51  CSNFLGASFS SSFINSSSNL SLLGKGLSLT FTSCQAPTNS NYALLSAAET
     101  LTFKNFSSIN FTGNQSTGLG GLIYGKDIVF QSIKDLIFTT NRVAYSPASV
     151  TTSATPAITT VTTGASALQP TDSLTVENIS QSIKFFGNLA NFGSAISSSP
     201  TAVVKFINNT ATMSFSHNFT SSGGGVIYGG SLLFENNSG CIIFTANSCV
    10  251  NSLKGVTSSS GTYALGSGGA ICIPTGTFFEL KNNQKCTFS YNGTPNDAGA
     301  IYAETCNIVG NQGALLLDSN TAARNNGAIC AKVLNIQGRG PIEFSRNRAE
     351  KGGAIFIGPS VGDPKQTST LTIASEGDI AFQGNMLNTH PGIRNAITVE
     401  AGGEIVSLSA QGGSRLVFDY PITHSLPTS PSNKDITINA NGASGSVFT
     451  SKGLSSTELL LPANTTTILL GTVKIASGEL KITDNAVNVN LGFATQSGSQ
    15  501  LTLGSGGTLG LATPTGAPAA VDFTIGKLAF DPFSFLKRDF VSASVNAGTK
     551  NVTLTGALVL DEHDVTDLYD MVSLQTPVAI PIAVFKGATV TKTGFPDGEI
     601  ATPSHYGYQG KWSYTWSRPL LIPAPDGGFP GGPSPSANTL YAVWNSDTLV
     651  RSTYILDPER YGEIVSNSLW ISFLGNQAFS DILQDVLLID HPGLSITAKA
     701  LGAYVEHTPR QGHEGFSGRY GGYQAALSMN YTDHTTLGLS FGQLYGKTNA
    20  751  NPYDSRCSEQ MYLLSFFGQF PIVTQKSEAL ISWKAAYGYS KNHLNTTYLR
     801  PDKAPKSQGG WHNNSYVILI SAEHPFLNWC LLTRPLAQAW DLSGFIISAEF
     851  LGGWQSKFTE TGDQLRSFSR GKGYNVSLPI GCSSQWFTPF KKAPSTLTIK
     901  LAYKPDIVRV NPHNIVTVVS NQESTSISGA NLRRHGLFVQ IHDVVDLTED
     951  TQAFNLNYTFD GKNGFTNHRV STGLKSTF*

```

25 A predicted signal peptide is highlighted.

The cp6737 nucleotide sequence <SEQ ID 36> is:

```

      1  ATGCCTCTTT CTTTCAAATC TTCATCTTTT TGTCTACTTG CCTGTTTATG
     51  TAGTGCAAGT TGC GCGTTTG CTGAGACTAG ACTCGGAGGG AACTTTGTTC
    101  CTCCAATTAC GAATCAGGCT GAAGAGATCT TACTCACTTC AGATTTTGTG
    151  TGTTCAAACT TCTTGGGGGC GAGTTTTCAT AGTTCCTTTA TCAATAGTTC
    201  CAGCAATCTC TCCTTATTAG GGAAGGGCCT TTCCTTAACG TTTACCTCTT
    251  GTCAAGCTCC TACAAATAGT AACTATGCGC TACTTTCTGC CGCAGAGACT
    301  CTGACCTTCA AGAATTTTTC TTCTATAAAC TTTACAGGGA ACCAATCGAC
    351  AGGACTTGGC GGCCTCATCT ACGGAAAAGA TATTGTTTTC CAATCTATCA
    401  AAGATTTGAT CTTCACTACG AACCGTGTTG CCTATTCTCC AGCATCTGTA
    451  ACTACGTCGG CAACTCCCGC AATCACTACA GTAACACAG GAGCCTCTGC
    501  TCTCCAACCT ACAGATCAC TCACTGTGCA AAACATATCC CAATCGATCA
    551  AGTTTTTTGG GAACCTTGCC AACTTCGGCT CTGCAATTAG CAGTTCTCCC
    601  ACGGCAGTCG TTAAATTCAT CAATAACACC GCTACCATGA GCTTCTCCCA
    651  TAACTTTACT TCGTCAGGAG GCGGCGTGAT TTATGGAGGA AGCTCTCTCC
    701  TTTTTGAAAA CAATTCGGA TGCATCATCT TCACCGCCAA CTCCTGTGTG
    751  AACAGCTTAA AAGGCGTCAC CCCTTCATCA GGAACCTATG CTTTAGGAAG
    801  TGGCGGAGCC ATCTGCATCC CTACGGGAAC TTTCGAATTA AAAACAATC
    851  AGGGGAAGTG CACCTTCTCT TATAATGGTA CACCAAATGA TCGGGGTGCG
    901  ATCTACGCCG AAACCTGCAA CATCGTAGGG AACCAGGGTG CCTTGCTCCT
    951  AGATAGCAAC ACTGCAGCGA GAAATGGCGG AGCCATCTGT GCTAAAGTGC
   1001  TCAATATTCA AGGACGCGGT CCTATTGAAT TCTCTAGAAA CCGCGCGGAG
   1051  AAGGGTGGAG CTATTTTCAT AGGCCCTCTT GTTGGAGACC CTGCGAAGCA
   1101  AACATCGACA CTTACGATTT TGGCTTCCGA AGGTGATATT GCGTTCCAAG
   1151  GAAACATGCT CAATCAAAA CCTGGAATCC GCAATGCCAT CACTGTAGAA
   1201  GCAGGGGGAG AGATTGTGTC TCTATCTGCA CAAGGAGGCT CACGTCTTGT
   1251  ATTTTATGAT CCCATTACAC ATAGCCTCCC AACCACAAGT CCGTCTAATA
   1301  AAGACATTAC AATCAACGCT AATGGCGCTT CAGGATCTGT AGTCTTTACA
   1351  AGTAAGGGAC TCTCCTCTAC AGAACTCCTG TTGCCTGCCA ACACGACAAC
   1401  TATACTTCTA GGAACAGTCA AGATCGCTAG TGGAGAACTG AAGATTACTG
   1451  ACAATGCCGT TGTCAATGTT CTTGGCTTCG CTAATCAGGG CTCAGGTGAG
   1501  CTTACCCTGG GCTCTGGAGG AACCTTAGGG CTGGCAACAC CCACGGGAGC
   1551  ACCTGCCGCT GTAGACTTTA CGATTGGAAA GTTAGCATTC GATCCTTTTT
   1601  CCTTCCTAAA AAGAGATTTT GTTTCAGCAT CAGTAAATGC AGGCACAAAA
   1651  AACGTCACCT TAACAGGAGC TCTGGTTCTT GATGAACATG ACGTTACAGA

```

1701 TCTTTATGAT ATGGTGTTCAT TACAAACTCC AGTAGCAATT CCTATCGCTG  
 1751 TTTTCAAAGG AGCAACCGTT ACTAAGACAG GATTTCCTGA TGGGGAGATT  
 1801 GCGACTCCAA GCCACTACGG CTACCAAGGA AAGTGGTCTT ACACATGGTC  
 1851 CCGTCCCCTG TTAATTCAG CTCTGTATGG AGGATTTCCT GGAGGTCCCT  
 1901 CTCCTAGCGC AAATACTCTC TATGCTGTAT GGAATTCAGA CACTCTCGTG  
 1951 CGTTCTACCT ATATCTTAGA TCCCGAGCGT TACGGAGAAA TTGTACAGCAA  
 2001 CAGCTTATGG ATTTCTTCT TAGGAAATCA GGCATTCTCT GATATTCTCC  
 2051 AAGATGTTCT TTTGATAGAT CATCCCGGT TGTCCATAAC CGCGAAAGCT  
 2101 TTAGGAGCCT ATGTCGAACA CACACCAAGA CAAGGACATG AGGGCTTTTC  
 2151 AGGTCGCTAT GGAGGCTACC AAGCTGCGCT ATCTATGAAC TACACGGACC  
 2201 ACATACGTT AGGACTTCT TTCGGGCAGC TTTATGGAAA AACTAACGCC  
 2251 AACCCTACG ATTACGTTG CTCAGAACA ATGTATTTAC TCTCGTCTT  
 2301 TGGTCAATTC CCTATCGTGA CTCAAAAGAG CGAGGCCTTA ATTTCTGGA  
 2351 AAGCAGCTTA TGGTTATTCC AAAAATCACC TAAATACCAC CTACCTCAGA  
 2401 CCTGACAAAG CTCCAAATC TCAAGGGCAA TGGCATAACA ATAGTTACTA  
 2451 TGTTCTTATT TCTGCAGAAC ATCCTTTCCT AAACGGTGT CTTCTTACAA  
 2501 GACCTCTGGC TCAAGCTTGG GATCTTTCAG GTTTTATTTT CGCAGAATTC  
 2551 CTAGTGGTT GGCAAAGTAA GTTCACAGAA ACTGGAGATC TGCAACGTAG  
 2601 CTTTAGTAGA GGTAAAGGT ACAATGTTT CCTACCGATA GGAATGTTT  
 2651 CTCAATGGTT CACACCATT AAGAAGGCTC CTTCTACACT GACCATCAAA  
 2701 CTTGCCTACA AGCCTGATAT CTATCGTGT CACCCCTACA ATATTGTGAC  
 2751 TGTCGTCTCA AACCAAGAGA GCACTTCGAT CTCAGGAGCA AATCTACGCC  
 2801 GCCACGGTT GTTTGTACAA ATCCATGATG TAGTAGATCT CACCGAGGAC  
 2851 ACTCAGGCCT TTCTAACTA TACCTTTGAC GGGAAAAATG GATTTACAAA  
 2901 CCACCGAGTG TCTACAGGAC TAAATCCAC ATTTTAA

The PSORT algorithm predicts an outer membrane location (0.940).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 18A. The recombinant protein was used to immunise mice, whose sera were used in an immunoblot analysis blot (Figure 18B) and for FACS analysis (Figure 18C). A his-tagged protein was also expressed.

The cp6737 protein was also identified in the 2D-PAGE experiment (Cpn0454) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6737 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 19

The following *C.pneumoniae* protein (PID 4377090) was expressed <SEQ ID 37; cp7090>:

1 MNIHSLWKLC TLLALLALPA CSLSPNYGWE DSCNTCHHTR RKKPSSFGFV  
 51 PLYTEEDFNP NPTFGEYDSK EEKQYKSSQV AAFRNITFAT DSYTIKGEEN  
 101 LAILTNLVHY MKKNPKATLY IEGHTDERGA ASYNLALGAR RANAIKEHLR  
 151 KQGISADRLS TISYGKEHPL NSGHNELAWQ QNRRTEFKIH AR\*

A predicted signal peptide is highlighted.

The cp7090 nucleotide sequence <SEQ ID 38> is:

1 ATGAATATAC ATTCCCTATG GAAACTTTGT ACTTTATTGG CTTTACTTGC  
 51 ATTGCCAGCA TGTAGCCTTT CCCCTAATTA TGGCTGGGAG GATTCTGTGA  
 101 ATACATGCCA TCATACAAGA CGAAAAAGC CTTCTTCTTT TGGCTTTGTT  
 151 CCTCTCTATA CCGAAGAGGA CTTTAACCTT AATTTTACCT TCGGTGAGTA  
 201 TGATTCCAAA GAAGAAAAAC AATACAAGTC AAGCCAAGTT GCAGCATTTT  
 251 GTAATATCAC CTTTGCTACA GACAGCTATA CAATTAAAGG TGAAGAGAAC  
 301 CTTGCGATTG TCACGAACCT GGTTCACACT ATGAAGAAAA ACCCGAAAGC  
 351 TACACTGTAC ATTGAAGGGC ATACTGACGA GCGTGGAGCT GCATCTCTATA  
 401 ACCTTGCTTT AGGAGCACGA CGAGCCAATG CGATTAAAGA GCATCTCCGA  
 451 AAGCAGGGAA TCTCTGCAGA TCGTCTATCT ACTATTTCTT ACGGAAAAGA

501 ACATCCTTTA AATTCGGGAC ACAACGAACT AGCATGGCAA CAAAATCGCC  
551 GTACAGAGTT TAAGATTCAT GCACGCTAA

The PSORT algorithm predicts an outer membrane location (0.790).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 19A.

5 A his-tagged protein was also expressed. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 19B) and for FACS analysis.

These experiments show that cp7090 is useful immunogen. These properties are not evident from the sequence alone.

### Example 20

10 The following *C.pneumoniae* protein (PID 4377091) was expressed <SEQ ID 39; cp7091>:

1 **MLRQLCFQVF FFCFASLVYA** EELEVVRSE HITLPIEVSC QTDTKDPKIQ  
51 KYLSSLTEIF CKDIALGDCL QPTAASKESS SPLAISLRH VPQLSVVLLQ  
101 SSKTPQTLCS FTISQNLSDV RQKIHHAADT VHVALTGIPG ISAGKIVFAL  
151 SSLGKDQKLK QGELWTTDID GKNLAPLTTE CSLSITPKWV GVGSNFPYLY  
15 201 VSYKGVFPIK FLGSLNTEG KKVLPKGNQ LMPTFSRKK LLAFFVADTYG  
251 NPDLFIQFIS LTSGFMRPR RLLNENFGTQ GNPSFNPEGS QLVFISNKG  
301 RPRLYIMSLD PEPQAPRLLT KKYRNSSCPA WSPDGKKIAF CSVIKGVRQI  
351 CIYDLSSGED YQLTTSPTNK ESPSWAIDSR HLVFSAGNAE ESELYLISLV  
401 TRKTNKIAIG VGEKRFPSWG AFPQPIKRT L\*

20 A predicted signal peptide is highlighted.

The cp7091 nucleotide sequence <SEQ ID 40> is:

1 ATGTTACGGC AACTATGCTT CCAAGTTTTT TTCTTTTGCT TCGCATCGCT  
51 AGTCTATGCT GAAGAATTAG AAGTTGTTGT CCGTTCGAA CATATCACGC  
101 TCCCTATTGA GGTCTCTTGC CAGACCGATA CGAAAGATCC AAAAATACAG  
25 151 AAATACCTCA GCTCGCTAAC GGAGATATTT TGCAAGGACA TTGCCCTAGG  
201 AGATTGTCTA CAACCCACAG CGGCTTCTAA AGAATCGTCA TCTCCTTTAG  
251 CAATATCTTT ACGGTTGCAT GTACCTCAGC TATCTGTAGT GCTTTTACAG  
301 TCTTCAAAAA CTCCTCAAAC CTTATGTTCT TTTACTATTT CTCAAATCT  
351 TTCTGTAGAT CGTCAAAAAA TCCATCACGC TGCTGATACA GTTCATTACG  
30 401 CCCTCACAGG GATTCCTGGA ATCAGTGCTG GGAAAATTGT TTTTGCTCTA  
451 AGTTCTTTAG GAAAAGATCA AAAGCTCAAG CAAGGAGAAT TATGGACTAC  
501 AGATTACGAT GGGAAAAACC TCGCCCTTTT AACCACAGAA TGTTCGCTCT  
551 CTATAACTCC AAAATGGGTG GGTGTGGGAT CAAATTTTCC CTATCTCTAT  
601 GTTTCGTATA AGTATGGTGT GCCTAAAATT TTTCTTGGTT CCCTAGAGAA  
35 651 CACTGAAGGT AAAAAAGTCC TTCCGTAAAA AGGCAACCAA CTCATGCCTA  
701 CGTTTCTTCC AAGAAAAAAG CTTTGTAGCTT TCGTTGCTGA TACGTATGGA  
751 AATCTGATT TATTTATTCA ACCGTTCTCA CTAACCTCAG GACCTATGGG  
801 TCGCCACAGT CGCCTCCTTA ATGAGAATTT CGGGACTCAA GGAATCCCT  
851 CCTTCAACCC TGAAGGATCC CAGCTTGCTT TTATATCGAA CAAAGACGGC  
40 901 CGTCCGCGTC TTTATATTAT GTCCCTCGAT CCTGAACCCC AAGCACCTCG  
951 CTTGCTGACA AAAAAATACA GAAATAGCAG TTGCCCTGCA TGGTCTCCAG  
1001 ATGGTAAAAA AATAGCCTTC TGCTCTGTAA TTAAAGGGGT GCGACAAATT  
1051 TGTATTTACG ATCTCTCCTC TGGAGAGGAT TACCAACTCA CTACGTCTCC  
1101 CACAAATAAA GAGAGTCCTT CTTGGGCTAT AGACAGCCGT CATCTTGTCT  
45 1151 TTAGTGCGGG GAATGCTGAA GAATCAGAGT TATATTTAAT CAGTCTAGTC  
1201 ACCAAAAAAA CTAACAAAAT TGCTATAGGA GTAGGAGAAA AACGGTTCCC  
1251 CTCTTGGGGT GCTTTCCTC AGCAACCGAT AAAGAGACA CTATGA

The PSORT algorithm predicts an inner membrane location (0.109).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 20A.

50 A his-tagged protein was also expressed. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 20B) and for FACS analysis.

These experiments show that cp7091 is a useful immunogen. These properties are not evident from the sequence alone.

### Example 21

The following *C.pneumoniae* protein (PID 4376260) was expressed <SEQ ID 41; cp6260>:

```

5      1  MRFSLCGFEL VFSFTLLSVF DTSLSATIS LTPEDSFHGD SQNAERSYNV
51     QAGDVYSLTG DVSISNVDNS ALNKACFNVT SGSVTFAGNH HGLYFNNISS
101    GTTKEGAVLC QDPQATARF SGFSTLSFIQ SPGDIKEQGC LYSKNALMLL
151    NNYVVFEQN QSKTRGGAIS GANVTIVGNY DSVSFYQNA TFGGAIHSSG
201    PLQIAVNQAE IRFAQNTAKN GSGGALYSDG DIDIDQNAVY LFRENEALTT
10     251  AIGKGGAVCC LPTSGSSTPV PIVTFSDNKQ LVFERNH SIM GGGAIYARKL
301    SISSGGPTLF INNISYANSQ NLGGAIAIDT GGEISLSAEK GTITFQGNRT
351    SLPFLNGIHL LQNAKFLKLQ ARNGYSIEFY DPITSEADGS TQLNINGDPK
401    NKEYTGILF SGEKSLANDP RDFKSTIPQN VNL SAGY LVI KEGAEVTVSK
451    FTQSPGSHLV LDLGTKLIAS KEDIAITGLA IDIDSLSSSS TAAVIKANTA
15     501    NKQISVTDSI ELISPTGNAY EDLRMRNSQT FPLLSLEPGA GGSVTVTAGD
551    FLPVSPHYGF QGNWKLAWTG TGNKVGEFFW DKINYKPRPE KEGNLVPNIL
601    WGNVDVRS L MQVQETHASS LQTDRLWID GIGNFFHVSA SEDNIRYRHN
651    SGGYVLSVNN EITPKHYTSM AFSQLFSRDK DYAVSNNEYR MYLGSYLYQY
701    TTSLGNIFRY ASRNPVNVVG ILSRRFLQNP LMIFHFLCAY GHATNDMKTD
20     751    YANFPMVKNS WRNNCWAIEC GGSMPLLVFE NGRLFQGAIP FMKLQLVYAY
801    QGDFKETAD GRRFSNGSLT SISVPLGIRF EKLALSQDVL YDFSFSYIPD
851    IFRKDPSCEA ALVISGDSWL VPAAHVSRHA FVGS GTGRYH FNDYTELLCR
901    GSIECRPHAR NYNINCGSKF RF*

```

A predicted signal peptide is highlighted.

25 The cp6260 nucleotide sequence <SEQ ID 42> is:

```

1      ATGCGATTTT CGCTCTGCGG ATTTCTCTCTA GTTTTTTCTT TTACATTGCT
51     CTCAGTCTTC GACACTTCTT TGAGTGCTAC TACGATTCTT TTAACCCAG
101    AAGATAGTTT TCATGGAGAT AGTCAGAATG CAGAACGTTT TTATAATGTT
151    CAAGCTGGGG ATGTCTATAG CCTTACTGGT GATGTCTCAA TATCTAACGT
30     201    CGATAACTCT GCATTAAATA AAGCCTGCTT CAATGTGACC TCAGGAAGTG
251    TGACGTTTCG AGGAAATCAT CATGGGTAT ATTTTAATAA TATTTCTCA
301    GGAAC TACAA AGGAAGGGGC TGTACTTTGT TGCCAAGATC CTCAAGCAAC
351    GGCACGTTTT TCTGGGTTCT CCACGCTCTC TTTTATTCAG AGCCCCGGAG
401    ATATTAAAGA ACAGGGATGT CTCTATTCAA AAAATGCACT TATGCTCTTA
35     451    AACAATTATG TAGTGCGTTT TGAACAAAC CAAAGTAAGA CTAAAGCGCG
501    AGCTATTAGT GGGGCGAATG TTA CTATAGT AGGCAACTAC GATTCCTGCT
551    CTTTCTATCA GAATGCAGCC ACTTTTGAG GTGCTATCCA TTCTTCAGGT
601    CCCCTACAGA TTGCAGTAAA TCAGGCAGAG ATAAGATTG CACAAAATAC
651    TGCCAAGAAT GGTCTTGAG GGGCTTTGTA CTCCGATGGT GATATTGATA
40     701    TTGATCAGAA TGCTTATGTT CTATTTGAG AAAATGAGGC ATTGACTACT
751    GCTATAGGTA AGGGAGGGGC TGTCTGTTGT CTTCCCACTT CAGGAAGTAG
801    TACTCCAGTT CCTATTGTGA CTTTCTCTGA CAATAAACAG TTAGTCTTTG
851    AAAGAAACCA TTCCATAATG GGTGGCGGAG CCATTTATGC TAGGAAACTT
901    AGCATCTCTT CAGGAGGTCC TACTCTATTT ATCAATAATA TATCATATGC
45     951    AAATTGCGAA AATTTAGGTG GAGCTATTGC CATTGATACT GGAGGGGAGA
1001   TCAGTTTATC AGCAGAGAAA GGAACAATTA CATTC CAAGG AAACCGGACG
1051   AGCTTACCGT TTTTGAATGG CATCCATCTT TTACAAAATG CTAAATTCCT
1101   GAAATTACAG GCGAGAAATG GATACTCTAT AGAATTTTAT GATCCTATTA
1151   CTTCTGAAGC AGATGGGTCT ACCCAATTGA ATATCAACGG AGATCCTAAA
50     1201   AATAAAGAGT ACACAGGGAC CATACTCTTT TCTGGAGAAA AGAGTCTAGC
1251   AAACGATCCT AGGGATTTTA AATCTACAAT CCCTCAGAAC GTCAACCTGT
1301   CTGCAGGATA CTTAGTTATT AAAGAGGGGG CCGAAGTCAC AGTTTCAAAA
1351   TTCACGCAGT CTCCAGGATC GCATTTAGTT TTAGATTTAG GAACCAACT
1401   GATAGCCTCT AAGGAAGACA TTGCCATCAC AGGCCTCGCG ATAGATATAG
55     1451   ATAGCTTAAG CTCATCCTCA ACAGCAGCTG TTATTAAAGC AAACACCGCA
1501   AATAAACAGA TATCCGTGAC GGA CTCTATA GAACTTATCT CGCCTACTGG
1551   CAATGCCAT GAAGATCTCA GAATGAGAAA TTCACAGACG TTCCCTCTGC
1601   TCTCTTTAGA GCCTGGAGCC GGGGGTAGTG TGA CTGTAA TGTGAGAT
1651   TTCCTACCGG TAAGTCCCCA TTATGGTTTT CAAGGCAATT GGAAATTAGC
60     1701   TTGGACAGGA ACTGGAACA AAGTTGGAGA ATTCTTCTGG GATAAAATAA

```

1751 ATTATAAGCC TAGACCTGAA AAAGAAGGAA ATTTAGTTCC TAATATCTTG  
 1801 TGGGGGAATG CTGTAGATGT CAGATCCTTA ATGCAGGTTT AAGAGACCCA  
 1851 TGCATCGAGC TTACAGACAG ATCGAGGGCT GTGGATCGAT GGAATTGGGA  
 1901 ATTTCTTCCA TGTATCTGCC TCCGAAGACA ATATAAGGTA CCGTCATAAC  
 1951 AGCGGTGGAT ATGTTCTATC TGTAATAAT GAGATCACAC CTAAGCACTA  
 2001 TACTTCGATG GCATTTTCCC AACTCTTTAG TAGAGACAAG GACTATCGCG  
 2051 TTTCCAACAA CGAATACAGA ATGTATTTAG GATCGTATCT CTATCAATAT  
 2101 ACAACCTCCC TAGGGAATAT TTTCCGTTAT GCTTCGCGTA ACCCTAATGT  
 2151 AAACGTCGGG ATTCTCTCAA GAAGGTTTCT TCAAAATCCT CTTATGATTT  
 2201 TTCATTTTTT GTGTGCTTAT GGTCATGCCA CCAATGATAT GAAAACAGAC  
 2251 TACGCAAATT TCCCTATGGT GAAAACAGC TGGAGAAACA ATTGTTGGGC  
 2301 TATAGAGTGC GGAGGGAGCA TGCCTCTATT GGTATTTGAG AACGGAAGAC  
 2351 TTTTCCAAGG TGCCATCCCA TTTATGAAAC TACAATTAGT TTATGCTTAT  
 2401 CAGGGAGATT TCAAAGAGAC GACTGCAGAT GGCCGTAGAT TTAGTAATGG  
 2451 GAGTTTAAAC TCGATTCTCG TACCTCTAGG CATACGCTTT GAGAAGCTGG  
 2501 CACTTTCTCA GGATGTACTC TATGACTTTA GTTTCTCCTA TATTCCTGAT  
 2551 ATTTTCCGTA AGGATCCCTC ATGTGAAGCT GCTCTGGTGA TTAGCCGAGA  
 2601 CTCCTGGCTT GTTCCGGCAG CACACGTATC AAGACATGCT TTTGTAGGGA  
 2651 GTGGAACGGG TCGGTATCAC TTTAACGACT ATACTGAGCT CTTATGTCCA  
 2701 GGAAGTATAG AATGCCGCCC CCATGCTAGG AATTATAATA TAAACTGTGG  
 2751 AAGCAAATTT CGTTTTTAG

The PSORT algorithm predicts an outer membrane location (0.921).

The protein was expressed in *E.coli* and purified both as a his-tag and GST-fusion product. The GST-fusion is shown in Figure 21A. This recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 21B) and for FACS analysis (Figure 21C).

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6260 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## Example 22

The following *C.pneumoniae* protein (PID 4376456) was expressed <SEQ ID 43; cp6456>:

1 MSSPVNNTPS APNIPIPAPT TPGIPTTKPR SSFIEKVIIIV AKYILFAIAA  
 51 TSGALGTILG LSGALTPGIG IALLVIFVVS MVLGLILKD SISGGEERRL  
 101 REEVSRTSE NQRLTVITTT LETEVKDLKA AKDQLTLEIE AFRNENGNLK  
 151 TTAEDLEEQV SKLSEQLAAL ERINQLIQAN AGDAQEISSE LKKLISGWDS  
 201 KVVEQINTSI QALKVLLGQE WVQEAQTHVK AMQEQIQALQ AEILGMHNQS  
 251 TALQKSVENL LVQDQALTRV VGELLESENK LSQACSALRQ EIEKLAQHET  
 301 SLQQRIDAML AQEQNLAEQV TALEKMKQEA QKAESEFIAC VRDRTFGRRE  
 351 TPPPTTPVVE GDESQEEDEG GTPPVSQPSS PVDRTGDDGQ \*

The cp6456 nucleotide sequence <SEQ ID 44> is:

1 ATGTCATCTC CTGTAAATAA CACACCCTCA GCACCAAACA TTCCAATACC  
 51 AGCGCCACG ACTCCAGGTA TTCCTACAAC AAAACCTCGT TCTAGTTTCA  
 101 TTGAAAAGGT TATCATTGTA GCTAAGTACA TACTATTGTC AATTGCAGCC  
 151 ACATCAGGAG CACTCGGAAC AATTCTAGGT CTATCTGGAG CGCTAACCCC  
 201 AGGAATAGGT ATTGCCCTTC TTGTATCTT CTTTGTCTTCT ATGGTGCTTT  
 251 TAGGTTTAAT CCTTAAAGAT TCTATAAGTG GAGGAGAAGA ACGCAGGCTC  
 301 AGAGAAGAGG TCTCTCGATT TACAAGTGAG AATCAACGGT TGACAGTCAT  
 351 AACCACAACA CTTGAGACTG AAGTAAAGGA TTTAAAAGCA GCTAAAGATC  
 401 AACTTACACT TGAAATCGAA GCATTTAGAA ATGAAAACGG TAATTTAAAA  
 451 ACAACTGCTG AGGACTTAGA AGAGCAGGTT TCTAAACTTA GCGAACAATT  
 501 AGAAGCACTA GAGCGAATTA ATCAACTTAT CCAAGCAAAC GCTGGAGATG  
 551 CTCAAGAAAT TTCGTCTGAA CTAAAGAAAT TAATAAGCGG TTGGGATTCC  
 601 AAAGTTGTTG AACAGATAAA TACTTCTATT CAAGCATTTA AAGTGTATT  
 651 GGGTCAAGAG TGGGTGCAAG AGGCTCAAAC ACACGTTAAA GCAATGCAAG  
 701 AGCAAATTCA AGCATTGCAA GCTGAAATTC TAGGAATGCA CAATCAATCT

```

751 ACAGCATTGC AAAAGTCAGT TGAGAATCTA TTAGTACAAG ATCAAGCTCT
801 AACAAAGAGTA GTAGGTGAGT TGTTAGAGTC TGAGAACAAG CTAAGCCAAG
851 CTGTTTCTGC GCTACGTCAA GAAATAGAAA AGTTGGCCCA ACATGAAACA
901 TCTTTGCAAC AACGTATTGA TGCATGCTA GCCCAAGAGC AAAATTTGGC
951 AGAGCAGGTC ACAGCCCTTG AAAAAATGAA ACAAGAAGCT CAGAAGGCTG
1001 AGTCCGAGTT CATTGCTTGT GTACGTGATC GAACCTTCGG ACGTCGTGAA
1051 ACACCTCCAC CAACAACACC TGTAAGTTGAA GGTGATGAAA GTCAAGAAGA
1101 AGACGAAGGA GGTACTCCCC CAGTATCACA ACCATCTTCA CCCGTAGATA
1151 GAGCAACAGG AGATGGTCAG TAA

```

10 The PSORT algorithm predicts inner membrane (0.127).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 22A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 22B) and for FACS analysis (Figure 22C). A his-tag protein was also expressed.

15 These experiments show that cp6456 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 23

The following *C.pneumoniae* protein (PID 4376729) was expressed <SEQ ID 45; cp6729>:

```

1  MKIPLHKLLI SSTLVTPILL SIATYGADAS LSPTDSFDGA GGSTFTPKST
51  ADANGTNYVL SGNVYINDAG KGTALTGCCF TETTGDLTFT GKGYSFSFNT
20  VDAGSNAGAA ASTTADKALT FTGFSNLSFI AAPGTTVASG KSTLSSAGAL
151 NLTDNGTILF SQNVSNNEANN NGGAITTKTL SISGNTSSIT FTSNSAKKLG
201 GAIYSSAAAS ISGNTGQLVF MNKGETGGG ALGFEASSI TQNSSLFFSG
251 NTATDAAGKG GAIYCEKTGE TPTLTISGNK SLTFAENSSV TQGGAIICAHG
301 LDLSAAGPTL PSNNRCGNTA AGKGAIAIA DSGSLSLSAN QGDITFLGNT
25  LTSTSAPTST RNAIYLSSA KITNLRAAQ QSIYFYDPIA SNTTGASDVL
401 TINQPDNSNP LDYSGTIVFS GEKLSADEAK AADNFTSILK QPLALASGTL
451 ALKGNVELDV NGFTQTEGST LLMQPGTKLK ADTEAISLTK LVVDLSALEG
501 NKSVSJETAG ANKTITLTSP LVFQDSSGNF YESHTINQAF TQPLVVFATA
551 TAASDIYIDA LLTSPVQTPE PHYGYQGHWE ATWADTSTAK SGTMTWVTTG
30  YNPNPERRAS VVPDSLWASF TDIRTLQQIM TSQANSIYQQ RGLWASGTAN
651 FFHKDKSGTN QAFRHKS YGY IVGGS AEDFS ENIFSVAF CQ LFGKDKDLFI
701 VENTSHNYLA SLYLQHRAFL GGLPMPSPGS ITDMLKDIPL ILNAQLSYSY
751 TKNDMDTRYT SYPEAQQSWT NNSGALELGG SLALYLPKEA PFFQGYFPFL
801 KFAQVYSRQQ NFKESGAEAR AFDDGDLVNC SIPVGIRLEK ISEDEKNNFE
35  ISLAYIGDVY RKNPRSR TSL MVSGASW TSL CKNLARQAFL ASAGSHLTLS
901 PHVELSGEAA YELRGS AHY NVDCGL RYSF *

```

A predicted signal peptide is highlighted.

The cp6729 nucleotide sequence <SEQ ID 46> is:

```

1  ATGAAAATAC CCTTGCACAA ACTCCTGATC TCTTCGACTC TTGTCACTCC
40  CATTCTATTG AGCATTGCAA CTTACGGAGC AGATGCTTCT TTATCCCCTA
101 CAGATAGCTT TGATGGAGCG GCGGCTCTA CATTTACTCC AAAATCTACA
151 GCAGATGCCA ATGGAACGAA CTATGTCTTA TCAGGAAATG TCTATATAAA
201 CGATGCTGGG AAAGGCACAG CATTACAGG CTGCTGCTTT ACAGAACTA
251 CGGGTGATCT GACATTTACT GGAAGGGAT ACTCATTTTC ATTCAACACG
45  GTAGATGCGG GTTCAATGCG AGGAGCTGCG GCAAGCACAA CTGCTGATAA
301 AGCCCTAACA TTCACAGGAT TTTCTAACCT TTCCTTCATT GCAGCTCCTG
351 GAACTACAGT TGCTTCAGGA AAAAGTACTT TAAGTTCTGC AGGAGCCTTA
401 AATCTTACCG ATAATGGAAC GATTCTCTTT AGCCAAAACG TCTCCAATGA
451 AGCTAATAAC AATGGCGGAG CGATCACCAC AAAAATCTTT TCTATTTCTG
50  GGAATACCTC TTCTATAACC TTCACTAGTA ATAGCGCAAA AAAATTAGGT
601 GGAGCGATCT ATAGCTCTGC GGCTGCAAGT ATTTCAAGAA ACACCGGCCA
651 GTTAGTCTTT ATGAATAATA AAGGAGAAAC TGGGGGTGGG GCTCTGGGCT
701 TTGAAGCCAG CTCCTCGATT ACTCAAAATA GCTCCCTTTT CTCTCTGGA
751 AACACTGCAA CAGATGCTGC AGGCAAGGGC GGGGCCATTT ATTGTGAAAA
55  AACAGGAGAG ACTCCTACTC TTAATATCTC TGGAAATAAA AGTCTGACCT
851 TCGCCGAGAA CTCTTCAGTA ACTCAAGGCG GAGCAATCTG TGCCCATGTT

```

5  
 10  
 15  
 20  
 25  
 30  
 35

```

901 CTAGATCTTT CCGCTGCTGG CCCTACCCTA TTTTCAAATA ATAGATGCGG
951 GAACACAGCT GCAGGCAAGG GCGGCGCTAT TGCAATTGCC GACTCTGGAT
1001 CTTTAACTCT CTCTGCAAAAT CAAGGAGACA TCACGTTCTT TGGCAACACT
1051 CTAACCTCAA CCTCCGCGCC AACATCGACA CGGAATGCTA TCTACCTGGG
1101 ATCGTCAGCA AAAATTACGA ACTTAAGGGC AGCCCAAGGC CAATCTATCT
1151 ATTTCTATGA TCCGATTGCA TCTAACACCA CAGGAGCTTC AGACGTTCTG
1201 ACCATCAACC AACCGGATAG CAACTCGCCT TTAGATTATT CAGGAACGAT
1251 TGTATTTTCT GGGGAAAAGC TCTCTGCAGA TGAAGCGAAA GCTGCTGATA
1301 ACTTCACATC TATATTAAAG CAACCATTGG CTCTAGCCTC TGGAACTTAA
1351 GCACTCAAAG GAAATGTCGA GTTAGATGTC AATGGTTTCA CACAGACTGA
1401 AGGCTCTACA CTCCTCATGC AACCAGGAAC AAAGCTCAAA GCAGATACTG
1451 AAGCTATCAG TCTTACCAA CTTGTCGTTG ATCTTTCTGC CTTAGAGGGA
1501 AATAAGAGTG TGTCCATTGA AACAGCAGGA GCCAACAAAA CTATAACTCT
1551 AACCTCTCCT CTTGTTTTC AAGATAGTAG CGGCAATTTT TATGAAAGCC
1601 ATACGATAAA CCAAGCCTTC ACGCAGCCTT TGGTGGTATT CACTGCTGCT
1651 ACTGCTGCTA GCGATATTTA TATCGATGCG CTTCTCACTT CTCCAGTACA
1701 AACTCCAGAA CCTCATTACG GGTATCAGGG ACATTGGGAA GCCACTTGGG
1751 CAGACACATC AACTGCAAAA TCAGGAACTA TGACTTGGGT AACTACGGGC
1801 TACAACCCTA ATCCTGAGCG TAGAGCTTCC GTAGTTCCCG ATTCAATTAG
1851 GGCATCCTTT ACTGACATTG CCACTCTACA GCAGATCATG ACATCTCAAG
1901 CGAATAGTAT CTATCAGCAA CGAGGACTCT GGGCATCAGG AACTGCGAAT
1951 TTCTTCCATA AGGATAAATC AGGAACAAAC CAAGCATTCG GACATAAAAG
2001 CTACGGCTAT ATTGTTGGAG GAAGTGCTGA AGATTTTCTT GAAAAATATCT
2051 TCAGTGTAGC TTTCTGCCAG CTCTTCGGTA AAGATAAAGA CCTGTTTATA
2101 GTTGAATAA CTTCTCATAA CTATTTAGCG TCGCTATACC TGCAACATCG
2151 AGCATTCCTA GGAGGACTTC CCATGCCCTC ATTTGGGAAG ATCACCAGCA
2201 TGCTGAAAGA TATTCTCTC ATTTTGAATG CCCAGCTAAG CTACAGCTAC
2251 ACTAAAAATG ATATGGATAC TCGCTATACT TCCTATCCTG AAGCTCAAGG
2301 CTCTTGGACC AATAACTCTG GGGCTCTAGA GCTCGGAGGA TCTCTGGCTC
2351 TATATCTCCC TAAAGAGCA CCGTCTTCC AGGGATATTT CCCCTTCTTA
2401 AAGTTCAGG CAGTCTACAG CCGCCAACAA AACTTTAAAG AGAGTGGCGC
2451 TGAAGCCCGT GCTTTTGATG ATGGAGACCT AGTGAACGTC TCTATCCCTG
2501 TCGGCATTCT GTTAGAAAAA ATCTCCGAAG ATGAAAAAAA TAATTTTCGAG
2551 ATTTCTCTAG CCTACATTGG TGATGTGTAT CGTAAAAATC CCCGTTTCGCG
2601 TACTTCTCTA ATGGTCAGTG GAGCCTCTTG GACTTCGCTA TGTAAAAACC
2651 TCGCAGGACA AGCCTTCTTA GCAAGTGCTG GAAGCCATCT GACTCTCTCC
2701 CCTCATGTAG AACTCTCTGG GGAAGCTGCT TATGAGCTTC GTGGCTCAGC
2751 ACACATCTAC AATGTAGATT GTGGGCTAAG ATACTCATTC TAG
  
```

The PSORT algorithm predicts outer membrane (0.927).

40 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 23A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 23B) and for FACS analysis (Figure 23C). A his-tag protein was also expressed.

The cp6729 protein was also identified in the 2D-PAGE experiment (Cpn0446) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

45 These experiments show that cp6729 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 24

The following *C.pneumoniae* protein (PID 4376849) was expressed <SEQ ID 47; cp6849>:

50

```

1  MSKLIRRVVT VLALTSMA SC FASGGIEAAV AESLITKIVA SAETKPAPVP
51  MTAKKVRVLR RNKQPEQKS RGAFCDFEYF PCEEGRCQPV EAQQESCYGR
101 LYSVKVND DC NVEICQSVPE YATVGSPYPI EILAIGKKDC VDVVITQQLP
151 CEAEFVSSDP ETTPTSDGKL VWKIDRLGAG DKCKITVWVK PLKEGCCFTA
201 ATVCACPELR SYTKCQPAI CIKQEGPDCA CLRCPCYKI EVVNTGSAIA
251 RNVTVDPNVP DGYSHASGQR VLSFNLGDMR PGDKKVFTVE FCPQRRGQIT
55 301 NVATVTYCGG HKCSANVTIV VNEPCVQVNI SGADWSYVCK PVEYSISVSN
351 PGDLVLHDVV IQDTLP SGVT VLEAPGGEIC CNKVVRRIKE MCPGETLQFK
  
```

```

401  LVVKAQVPGR FTNQVAVTSE SNCGTCTSCA ETTTHWKGLA ATHMCVLDTN
451  DPICVGENTV YRICVTNRGS AEDTNVSLIL KFSKELQPIA SSGPTKGTIS
501  GNTVVPDALP KLGSKESVEF SVTLKGIAPG DARGEAILSS DTLTSPVSDT
551  ENTHVY*

```

5 A predicted signal peptide is highlighted.

The cp6849 nucleotide sequence <SEQ ID 48> is:

```

1  ATGTCCAAAC TCATCAGACG AGTAGTTACG GTCCTTGCGC TAACGAGTAT
51  GGCAGAGTTGC TTTGCCAGCG GGGGTATAGA GGCCGCTGTA GCAGAGTCTC
101 TGATTACTAA GATCGTCGCT AGTGCAGGAA CAAAGCCAGC ACCTGTTCCT
151 ATGACAGCGA AGAAGGTTAG ACTTGTCGGT AGAAATAAAC AACCAGTTGA
201 ACAAAAAAGC CGTGGTGCTT TTTGTGATAA AGAATTTTAT CCCTGTGAAG
251 AGGGACGATG TCAACCTGTA GAGGCTCAGC AAGAGTCTTG CTACGGAAGA
301 TTGTATTCTG TAAAAGTAAA CGATGATTGC AACGTAGAAA TTTGCCAGTC
351 CGTTCCAGAA TACGCTACTG TAGGATCTCC TTACCCTATT GAAATCCTTG
401 CTATAGGCAA AAAAGATTGT GTTGATGTTG TGATTACACA ACAGCTACCT
451 TGCGAAGCTG AATTCGTAA GAGTGATCCA GAAACAATC CTACAAGTGA
501 TGGGAAATTA GTCTGGAAAA TCGATCGCCT GGGTGCAGGA GATAAATGCA
551 AAATTACTGT ATGGGTAAAA CCTCTTAAAG AAGGTTGCTG CTTACAGCT
601 GCTACTGTAT GTGCTTGCCC AGAGCTCCGT TCTTATACTA AATGCGGTCA
651 ACCAGCCATT TGTATTAAAG AAGAAGGACC TGACTGTGCT TGCCTAAGAT
701 GCCCTGTATG CTACAAAATC GAAGTAGTGA ACACAGGATC TGCTATTGCC
751 CGTAACGTAA CTGTAGATAA TCCTGTTCCC GATGGCTATT CTCATGCATC
801 TGGTCAAAGA GTTCTCTCTT TTAACCTAGG AGACATGAGA CCTGGCGATA
851 AAAAGGTATT TACAGTTGAG TTCTGCCCTC AAAGAAGAGG TCAAATCACT
901 AACGTTGCTA CTGTAACCTA CTGCGGTGGA CACAAATGTT CTGCAAATGT
951 AACTACAGTT GTTAATGAGC CTTGTGTACA AGTAAATATC TCTGGTGCTG
1001 ATTGGTCTTA CGTATGTAAA CCTGTGGAGT ACTCTATCTC AGTATCGAAT
1051 CCTGGAGACT TGGTTCTTCA TGATGTCGTG ATCCAAGATA CACTCCCCTC
1101 TGGTGTTACA GTACTCGAAG CTCCTGGTGG AGAGATCTGC TGTAAATAAG
1151 TTGTTTGCGG TATTAAAGAA ATGTGCCAG GAGAAACCCT CCAGTTTAAA
1201 CTTGTAGTGA AAGCTCAAGT TCCTGGAAGA TTCACAAATC AAGTTGCAGT
1251 AACTAGTGAG TCTAACTGCG GAACATGTAC ATCTTGCGCA GAAACAACAA
1301 CACATTGGAA AGGTCTTGCA GCTACCCATA TGTGCGTATT AGACACAAAT
1351 GATCCTATCT GTGTAGGAGA AAATACTGTC TATCGTATCT GTGTAACATA
1401 CCGTGGTTCT GCTGAAGATA CTAACGTATC TTTAATCTTG AAGTTCTCAA
1451 AAGAACTTCA GCCAATAGCT TCTTCAGGTC CAACTAAAGG AACGATTTCA
1501 GGTAAATACG TTGTTTTTCA CGCTTTACCT AAACCTCGGT CTAAGGAATC
1551 TGTAAGATTT TCTGTTACCT TGAAAGGTAT TGCTCCCGGA GATGCTCGCG
1601 GCGAAGCTAT TCTTTCTTCT GATACACTGA CTTACCAGT ATCAGACACA
1651 GAAAATACCC ACGTGTATTA A

```

The PSORT algorithm predicts periplasmic space (0.93).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 24A, and also as a his-tag protein. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 24B) and for FACS analysis (Figure 24C).

45 The cp6849 protein was also identified in the 2D-PAGE experiment (Cpn0557).

These experiments show that cp6849 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 25

The following *C.pneumoniae* protein (PID 4376273) was expressed <SEQ ID 49; cp6273>:

```

50  1  MGLFHLTLFG LLLCSLPISL VAKFPESVGH KILYISTQST QQALATYLEA
51  51  LDAYGDHDFV VLKRGEDYL KQSIHSSDPQ TRKSTIIGAG LAGSSEALDV
101 101  LSQAMETADP LQQLLVLSAV SGHLGKTSDD LLFKALASPY PVIRLEAAYR
151 151  LANLKNKTVI DHLHSFIHKL PEEIQCLSAA IFLRLETEES DAYIRDLLAA
201 201  KKSARSATA LQIGEYQQR FLPTLRNLLT SASPQDQEI LYALGKLDG

```



```

251 QSYYNIKKQL QKPDVDVTLA AAQALIALGK EEDALPVIKK QALEERPRAL
301 YALRHLPSEI GIPIALPIFL KTKNSEAKLN VALALLELG C DTPKILLEYIT
351 ERLVQPHYNE TLALSFSKGR TLQNWKR VNI IVPQDPQERE RLLSTTRGLE
401 EQILTFLFRL PKEAYLPCY KLLASQKTQL ATTAISFLSH TSHQALDLL
451 FQAAKLPGEPI IIRAYADLAI YNLTKDPEKK RSLHDYAKKL IQETLLFVDT
501 ENQRPHPSMP YLRYQVTPES RTKLMLDILE TLATSKSSED IRLLIQLMTE
551 GDAKNFPVLA GLLIKIVE*

```

A predicted signal peptide is highlighted.

The cp6273 nucleotide sequence <SEQ ID 50> is:

```

10      1  ATGGGACTAT TCCATCTAAC TCTCTTTGGA CTTTATTGT GTAGTCTTCC
      51  CATTTCTCTT GTTGCTAAAT TCCCTGAGTC TGTAGGTCAT AAGATCCTTT
     101  ATATAAGTAC GCAATCTACA CAGCAGGCCT TAGCAACATA TCTGGAAGCT
     151  CTAGATGCCT ACGGTGATCA TGACTTCTTC GTTTTAAGAA AAATCGGAGA
     201  AGACTATCTC AAGCAAAGCA TCCACTCCTC AGATCCGCAA ACTAGAAAAA
     251  GCACCATCAT TGGAGCAGGC CTGGCGGGAT CTTCAGAAGC CTTGGACGTG
     301  CTCTCCCAAG CTATGGAAC TGCAGACCCC CTGCAGCAGC TACTGGTTTT
     351  ATCGGCAGTC TCAGGACATC TTGGGAAAAC TTCTGACGAC TTACTGTTTA
     401  AAGCTTTAGC ATCTCCCTAT CCTGTCATCC GCTTAGAAGC CGCCTATAGA
     451  CTTGCTAATT TGAAGAACAC TAAAGTCATT GATCATCTAC ATTCTTTCAT
     501  TCATAAGCTT CCCGAAGAAA TCCAATGCCT ATCTGCGGCA ATATTCCCTAC
     551  GCTTGGAGAC TGAAGAATCT GATGCTTATA TTCGGGATCT CTTAGCTGCC
     601  AAGAAAAGCG CGATTCCGAG TGCCACAGCT TTGCAGATCG GAGAATACCA
     651  ACAAAAACGC TTTCTTCCGA CACTTAGGAA TTTGCTAACG AGTGCCTCTC
     701  CTCAAGATCA AGAAGCTATT CTTTATGCTT TAGGGAAGCT TAAGGATGGT
     751  CAGAGCTACT ACAATATAAA AAAGCAATTG CAGAAGCCTG ATGTGGATGT
     801  CACTTTAGCA GCAGCTCAAG CTTTAATTGC TTTGGGAAA GAAGAGGACG
     851  CTCTTCCCGT GATAAAAAG CAAGCACTTG AGGAGCGGCC TCGAGCCCTG
     901  TATGCCTTAC GGCATCTACC CTCTGAGATA GGGATTCCGA TTGCCCTGCC
     951  GATATTCTTA AAAACTAAGA ACAGCGAAGC CAAGTTGAAT GTAGCTTTAG
    1001  CTCTCTTAGA GTTAGGGTGT GACACCCCTA AACTACTGGA ATACATTACC
    1051  GAAAGGCTTG TCCAACCACA TTATAATGAG ACTCTAGCCT TGAGTTTCTC
    1101  TAAGGGGCGT ACTTTACAAA ATTGGAAGCG GGTGAACATC ATAGTCCCTC
    1151  AAGATCCCCA GGAGAGGGAA AGGTTGCTCT CCACAACCCG AGGTCTTGAA
    1201  GAGCAGATCC TTACGTTTCT CTTCCGCCTA CCTAAAGAAG CTTACCTCCC
    1251  CTGTATTAT AAGCTTTTGG CGAGTCAGAA AACTCAGCTT GCCACTACTG
    1301  CGATTTCTTT TTAAAGTCAC ACCTCACATC AGGAAGCCTT AGATCTACTT
    1351  TTCCAAGCTG CGAAGCTTCC TGGAGAACCT ATCATCCGCG CCTATGCAGA
    1401  TCTTGCTATT TATAATCTCA CCAAAGATCC TGAAAAAAA CGTTCTCTCC
    1451  ATGATTATGC AAAAAAGCTA ATTCAGGAAA CCTTGTTATT TGTGGACACG
    1501  GAAAACCAAA GACCCCATCC CAGCATGCCC TATCTACGTT ATCAGGTCAC
    1551  CCCGAAAGC CGTACGAAGC TCATGTTGGA TATTCTAGAG ACACTAGCCA
    1601  CCTCGAAGTC TTCCGAAGAT ATCCGTTTAT TGATACAACT GATGACGGAA
    1651  GGAGATGCAA AAAATTTCCC AGTCCTTGCA GGCTTACTCA TAAAAATTGT
    1701  GGAGTAA

```

45 The PSORT algorithm predicts a periplasmic location (0.922).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 25A. The recombinant GST-fusion was used to immunise mice, whose sera were used in a Western blot (Figure 25B) and for FACS analysis (Figure 25C).

50 This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6273 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 26

The following *C.pneumoniae* protein (PID 4376735) was expressed <SEQ ID 51; cp6735>:

-68-

1 **MTILRNFLTC** **SALFLALPAA** AQQVYLHESD GYNGAINNKS LEPKITCYPE  
 51 GTSYIFLDDV RISNVKHDQE DAGVFINRSG NLFFMGNRCN FTFHNLMTG  
 101 FGAAISNRVG DTTLTLSNFS YLAFTSAPLL PQGGGAIYSL GSVMIENSEE  
 151 VTFCGNYSSW SGAAIYTPYL LGSKASRPSV NLSGNRYLVF RDNVSQGYGG  
 5 201 AISTHNLTLT TRGPSCFENN HAYHDVNSNG GAIATAPGGS ISISVKSDDL  
 251 IFKGNTASQD GNTIHNSIHL QSGAQFKNLR AVSESGVYFY DPISHSESHK  
 301 ITDLVINAPE GKETYEGTIS FSGLCLDDHE VCAENLTSTI LQDVTLAGGT  
 351 LSLSDGVTLQ LHSFKQEASS TLTMSPGTTL LCSGDARVQN LHILIEDTDN  
 401 FVPVRIRAEK KDALVSLEKL KVAFEAYWSV YDFPQFKEAF TIPLLELLGP  
 10 451 SFDSLILGET TLERTQVTTE NDAVRGFWSL SWEYPPSLD KDRRITPTTK  
 501 TVFLTWNPEI TSTP\*

A predicted signal peptide is highlighted.

The cp6735 nucleotide sequence <SEQ ID 52> is:

1 ATGACCATAC TTCGAAATTT TCTTACCTGC TCGGCTTTAT TCCTCGCTCT  
 15 51 CCCTGCAGCA GCACAAGTTG TATATCTTCA TGAAAGTGAT GGTATAACG  
 101 GTGCTATCAA TAATAAAGC TTAGAACCTA AAATTACCTG TTATCCAGAA  
 151 GGAACCTTCT ACATCTTTCT AGATGACGTG AGGATTCCA ACGTTAAGCA  
 201 TGATCAAGAA GATGCTGGGG TTTTATAAAA TCGATCTGGG AATCTTTTTT  
 251 TCATGGGCAA CCGTTGCAAC TTCACCTTTC ACAACCTTAT GACCGAGGGT  
 20 301 TTTGGCGCTG CCATTTGCAA CCGCGTTGGA GACACCACTC TCACTCTCTC  
 351 TAATTTTTCT TACTTAGCGT TCACCTCAGC ACCTCTACTA CCTCAAGGAC  
 401 AAGGAGCGAT TTATAGTCTT GGTTCCTGTA TGATCGAAAA TAGTGAGGAA  
 451 GTGACTTTCT GTGGGAATA CTCTTCGTGG AGTGGAGCTG CGATTTATAC  
 501 TCCCTACCTT TTAGGTTCTA AGGCGAGTCG TCCTTCAGTA AATCTCAGCG  
 25 551 GGAACCGCTA CCTGGTGTCT AGAGACAATG TGAGCCAAGG TTATGGCGGC  
 601 GCCATATCTA CCCACAATCT CACACTCACG ACTCGAGGAC CTTCTGTGTT  
 651 TGAAAATAAT CATGCTTATC ATGACGTGAA TAGTAATGGA GGAGCCATTG  
 701 CCATTGCTCC TGGAGGATCG ATCTCTATAT CCGTGAAAAG CGGAGATCTC  
 751 ATCTTCAAAG GAAATACAGC ATCACAAGAC GGAAATACAA TACACAACCT  
 30 801 CATCATCTG CAATCTGGAG CACAGTTTAA GAACCTACGT GCTGTTTTCAG  
 851 AATCCGGAGT TTATTTCTAT GATCCTATAA GCCATAGCGA GTCGCATAAA  
 901 ATTACAGATC TTGTAATCAA TGCTCCTGAA GGAAAGGAAA CTTATGAAGG  
 951 AACAATTAGC TTCTCAGGAC TATGCCTGGA TGATCATGAA GTTTGTGCGG  
 1001 AAAATCTTAC TTCCACAATC CTACAAGATG TCACATTAGC AGGAGGAACT  
 35 1051 CTCTCTCTAT CGGATGGGGT TACCTTGCAA CTGCATTCTT TTAAGCAGGA  
 1101 AGCAAGCTCT ACGCTTACTA TGTCTCCAGG AACCCTCTG CTCTGCTCAG  
 1151 GAGATGCTCG GGTTCAGAA CTGCACATCC TGATTGAAGA TACCGACAAC  
 1201 TTTGTTCTCG TAAGGATTCG CGCCGAGGAC AAGGATGCTC TTGTCTCATT  
 1251 AGAAAACTT AAAGTTGCCT TTGAGGCTTA TTGGTCCGTC TATGACTTTC  
 40 1301 CTAATTTTAA GGAAGCCTTT ACGATTCTCT TTCTTGAAC TCTAGGGCCT  
 1351 TCTTTTGACA GTCTTCTCCT AGGGGAGACC ACTTTGGAGA GAACCCAAGT  
 1401 CACAACAGAG AATGACGCCG TTCGAGGTTT CTGGTCCCTA AGCTGGGAAG  
 1451 AGTACCCCCC TTCTCTGGAT AAAGACAGAA GGATCACACC AACTAAGAAA  
 1501 ACTGTTTTCC TCACTTGGA TCCTGAGATC ACTTCTACGC CATAA

45 The PSORT algorithm predicts an outer membrane location (0.922).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 26A. The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 26B).

50 These experiments show that cp6735 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 27

The following *C.pneumoniae* protein (PID 4376784) was expressed <SEQ ID 53; cp6784>:

1 **MNRRKARWV** **ALFAMTALIS** **VGCCPWSQAK** SRCSDKYIP VVNRLLLEVCG  
 51 LPEAENVEDL IESSAWVLT PEERFSGELV SICQVKDEHA FYNDLSLLHM  
 55 101 TQAVPSYSAT YDCAVVFGGP LPALRQRLDF LVREWQRGVR FKKIVFLCGE  
 151 RGRYQSIEEQ EHFDSRYNP FPTEENWESG NRVTSPSEEE IAKFVWMQML

-69-

201 LPRAWRDSTS GVRVTFLAK PEENRVVANR KDTLLLLFRSY QRAFPGRVLF  
 251 VSSQFFIGLD ACRVGFQFFKG ESYDLAGPGF AQGVLYKHYWA PRICLHTLAE  
 301 WLKETNGCLN ISEGCFG\*

A predicted signal peptide is highlighted.

5 The cp6784 nucleotide sequence <SEQ ID 54> is:

1 ATGAATAGAA GAAAAGCAAG ATGGGTAGTG GCATTGTTTCG CAATGACGGC  
 51 GCTCATTCTT GTTGGGTGTT GTCCTTGGTC ACAAGCGAAA TCAAGATGTT  
 101 CTATTGATAA GTATATTCCCT TAGTCAATC GTTTACTAGA AGTTTGTGGA  
 151 CTTCCTGAAG CTGAGAATGT TGAGGATTTA ATCGAGTCCT CGTCTGCTTG  
 10 201 GGTACTGACT CCTGAAGAAC GTTTTCTGAG AGAGTTAGTC TCTATCTGTC  
 251 AGGTTAAAGA TGAGCATGCT TTCTATAACG ATTTGTCTTT ATTACATATG  
 301 ACTCAGGCTG TGCCTTCGTA TTCTGCAACG TATGATTGTG CTGTAGTTTTT  
 351 TGGCGGGCCT TTGCCAGCGC TACGTCAGCG CTTAGATTTT TTGGTGCAGAG  
 401 AGTGGCAGCG TGGCGTGCAG TTTAAGAAAA TCGTTTTTCT ATGTGGAGAG  
 15 451 CGAGGGCGCT ATCAGTCTAT TGAAGAACAA GAGCATTCTT TTGATTCTCG  
 501 GTACAATCCT TTCCCTACTG AAGAGAACTG GGAATCTGGT AACCGAGTTA  
 551 CTCCTCTTTC TGAAGAAGAG ATTGCCAAAT TTGTTTGGAT GCAAATGCTT  
 601 TTACCTAGAG CATGGCGAGA TAGTACTTCA GGAGTCAGAG TGACATTTCT  
 651 TCTAGCAAAG CCAGAGGAAA ATCGTGTGGT TCGCAATCGT AAGGACACCT  
 20 701 TACTTTTATT CCGTTCTTAT CAAGAAGCGT TTCCGGGACG CGTGTTATTT  
 751 GTAAGTAGTC AACCTTTTAT CGGTTTAGAT GCTTGCAGGG TCGGGCAGTT  
 801 TTTCAAAGGG GAAAGCTATG ATCTTGCTGG ACCTGGATT TCTCAAGGAG  
 851 TCTTGAAGTA TCATTGGGCT CCAAGGATT GTCTACATAC TTTAGCGGAA  
 901 TGGTTAAAGG AAACGAACGG CTGCTTAAAT ATTCAGAGG GTTGTTTTGG  
 25 951 ATGA

The PSORT algorithm predicts a periplasmic location (0.894).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 27A. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 27B). The GST-fusion product was used for FACS analysis (Figure 27C).

30 The cp6784 protein was also identified in the 2D-PAGE experiment (Cpn0498).

These experiments show that cp6784 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 28

The following *C.pneumoniae* protein (PID 4376960) was expressed <SEQ ID 55; cp6960>:

35 1 MNRRWNLVLA TVALALSVAS CDVRSKDKDK DQGSLEYKDK NKDTNDIELS  
 51 DNQKLSRTFG HLLARQLRKS EDMFFDIAEV AKGLQALVC KSAPLTETET  
 101 EEKMAEVQKL VFEKSKENL SLAEKFLKEN SKNAGVVEVQ PSKLQYKIIK  
 151 EGAGKAISGK PSALLHYKGS FINGQVFSSS EGNNEPILLP LGQTIPGFAL  
 201 GMQGMKEGET RVLYIHPDLA YGTAGQLPPN SLLIFEINLI QASADEVAAY  
 40 251 PQEGNQGE\*

A predicted signal peptide is highlighted.

The cp6960 nucleotide sequence <SEQ ID 56> is:

1 ATGAACAGAC GGTGGAATTT AGTTTATGCA ACAGTAGCTC TGGCACTCTC  
 51 CGTCGCTTCT TGTGACGTAC GGTCTAAGGA TAAAGACAAG GATCAGGGGT  
 45 101 CGTTAGTGGA ATATAAAGAT AACAAAGATA CCAATGACAT AGAATTATCC  
 151 GATAATCAAA AGTTATCCAG AACATTTGGT CATTATATTAG CACGCCAATT  
 201 ACGCAAGTCA GAAGATATGT TTTTGTATAT TGCAGAAGTG GCTAAGGGGT  
 251 TGCAGGCGGA ATTGGTTTGT AAAAGTGCTC CTTTAACAGA AACAGAGTAT  
 301 GAAGAAAAAA TGGCTGAAGT ACAGAAGTTG GTTTTGAAGA AAAAATCAAA  
 50 351 AGAAAAATCT TCATTGGCAG AAAAATCTT AAAAGAAAAA AGCAAGAACG  
 401 CTGGTGTGTG TGAAGTGCAA CCAAGTAAAT TGCAATACAA AATTATTAAA

451 GAAGGTGCAG GGAAAGCAAT TTCAGGTAAA CCTTCAGCTC TATTGCACTA  
 501 CAAGGGTTCC TTCATCAATG GCCAAGTATT TAGCAGTTCA GAAGGCAACA  
 551 ATGAGCCTAT CTGTCTTCCT CTAGGCCAAA CAATTCCCTGG TTTTGCTTTA  
 601 GGTATGCAGG GCATGAAAGA AGGAGAAACT CGAGTTCTCT ACATCCATCC  
 651 TGATCTTGCT TACGGAACCG CAGGACAACT TCCTCCAAAC TCTTTATTAA  
 701 TTTTGAAT TAACCTGATT CAGGCTTCAG CAGATGAAGT TGCTGCTGTA  
 751 CCCCAGAAG GAAATCAAGG TGAATGA

The PSORT algorithm predicts periplasmic space location (0.930).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 28A. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 28B) and for FACS analysis (Figure 28C).

The cp6960 protein was also identified in the 2D-PAGE experiment.

These experiments show that cp6960 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## Example 29

The following *C.pneumoniae* protein (PID 4376968) was expressed <SEQ ID 57; cp6968>:

1 **MKFLLYVPLL LVLVSTG**CDA KPVSEFPFSG KLSTQRFEPQ HSAEEYFSQG  
 51 QEFLKKGNFR KALLCFGIIT HHFPRDILRN QAQYLIGVCY FTQDHPDLAD  
 101 KAFASYLQLP DAEYSEELFQ MKYAIQRFA QGKRKRICRL EGFPKLMNAD  
 151 EDALRIYDEI LTAFFPSKDLG AQALYSKAAL LIVKNDLTEA TKTLKKLTLQ  
 201 FPLHILSSEA FVRLSEIYLQ QAKKEPHNLQ YLHFAKLNEE AMKKQHPNHP  
 251 LNEVVSANVG AMREHYARGL YATGRFYEKK KKAEEANIYY RTAITNYPDT  
 301 LLVAKCQKRL DRISKHTS\*

A predicted signal peptide is highlighted.

The cp6968 nucleotide sequence <SEQ ID 58> is:

1 ATGAAATTTC TATTATACGT TCCACTTCTT CTGTGTTCTCG TATCTACGGG  
 51 GTGCGATGCA AAACCTGTTT CTTTGTAGCC CTTTTCAGGA AAGCTTTCCA  
 101 CCCAGCGTTT TGAGCCTCAG CACTCTGCTG AAGAATATTT TTCTCAGGGA  
 151 CAGGAATTCT TAAAAAAGG AAATTCAGA AAAGCTTTAC TATGCTTTGG  
 201 AATCATTACG CATCACTTCC CTAGGGACAT CTTGCGTAAT CAAGCACAGT  
 251 ATCTTATAGG AGTCTGTTAC TTCACGCAGG ATCACCAGGA TTTAGCAGAC  
 301 AAGGCATTTG CATCTTACTT ACAACTTCC TATGCGGAGT ACTCTGAAGA  
 351 GTTGTTCAG ATGAAATATG CGATTGCTCA AAGATTGCT CAAGGGAAGC  
 401 GTAAACGGAT TTGTGCGATTA GAGGGCTTCC CAAAATAAT GAATGCTGAT  
 451 GAAGATGCGC TACGCATTTA TGACGAGATT CTAACAGCGT TTCCTAGTAA  
 501 AGACTTAGGA GCTCAGGCC TCTATAGTAA AGCTGCGTTA CTTATTGTAA  
 551 AAAACGATCT TACAGAAGCC ACCAAAACCT TAAAAAACT CACGTTACAA  
 601 TTCTCTCTAC ATATTTTATC TTCAGAGGCC TTTGTACGTT TATCGGAAAT  
 651 CTATTTACAG CAAGCTAAGA AAGAGCCTCA CAATCTTCAA TATCTTCATT  
 701 TTGCAAAAGCT TAATGAAGAG GCAATGAAAA AGCAGCATCC TAACCATCCT  
 751 CTGAATGAGG TTGTTTCTGC TAATGTTGGA GCTATGCGGG AACATTATGC  
 801 TCGAGGTTTG TATGCCACAG GTCGTTTCTA TGAGAAGAAG AAAAAAGCCG  
 851 AGGTGCGGAA TATCTATTAC CGCACTGCGA TTACAAACTA CCCAGACACT  
 901 TTATTAGTGG CTAAATGTCA AAAGCGTCTA GATAGAATAT CTAAGCATAC  
 951 TTCTTAA

The PSORT algorithm predicts an inner membrane location (0.790).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 29A. The recombinant GST-fusion was used to immunise mice, whose sera were used in a Western blot (Figure 29B) and for FACS analysis (Figure 29C).

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6968 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## 5 Example 30

The following *C.pneumoniae* protein (PID 4376998) was expressed <SEQ ID 59; cp6998>:

```

1  MKKLLKSALL SAAFAGSVGS LQALPVGNPS DPSLLIDGTI WEGAAGDPCD
51 PCATWCD AIS LRAGFYGDYV FDRILKVDAP KTF SMGAKPT GSAAANYTTA
101 VDRPNPAYNK HLHDAEWFTN AGFIALNIWD RFDVFC TLGA SNGYIRGNST
151 AFNLVGLFGV KGTTVNANEL PNVLSNGVV ELYTDT SFSW SVGARGALWE
201 CGCATLGAEF QYAQSKPKVE ELNVICNVSQ FSVNKP GYK GVAFFLP TDA
251 GVATATG TKS ATINYHEWQV GASLSYRLNS LVPYIGV QWS RATFDAD NIR
301 IAQPKLP TAV LNLTAWNPSL LGNATALSTT DSFSDFM QIV SCQINKFKSR
351 KACGVTVGAT LVDADKWSLT AEARLINERA AHVSGQFRF*

```

15 A predicted signal peptide is highlighted.

The cp6998 nucleotide sequence <SEQ ID 60> is:

```

1  ATGAAAAAC TCTTAAAGTC GGC GTTATTA TCCGCCGCAT TTGCTGGTTC
51 TGT TTGGCTCC TTACAAGCCT TGCCTGTAGG GAACCTTCT GATCCAAGCT
101 TATTAATTGA TGGTACAATA TGGGAAGGTG CTGCAGGAGA TCCTTGCGAT
201 CCTTGCGCTA CTTGGTGCGA CGCTATTAGC TTACGTGCTG GATTTTACGG
251 AGACTATGTT TTCGACCGTA TCTTAAAGT AGATGCACCT AAAACATTTT
301 CTATGGGAGC CAAGCCTACT GGATCCGCTG CTGCAAACTA TACTACTGCC
351 GTAGATAGAC CTAACCCGGC CTACAATAAG CATTTACACG ATGCAGAGTG
401 GTTCACTAAT GCAGGCTTCA TTGCCTTAAA CATTTGGGAT CGCTTTGATG
451 TTTTCTGTAC TTTAGGAGCT TCTAATGGTT ACATTAGAGG AAACCTCATA
501 CCGTTCAATC TCGTTGTTTT ATTTCGAGTT AAAGGTACTA CTGTAATGTC
551 AAATGA ACTA CCAAACGTTT CTTTAAGTAA CGGAGTTGTT GAACTTTACA
601 CAGACACCTC TTTCTCTTGG AGCGTAGGCG CTCGTGGAGC CTTATGGGAA
651 TGCGGTGTG CAACCTTGGG AGCTGAATTC CAATATGCAC AGTCCAAACC
701 TAAAGTTGAA GAACTTAATG TGATCTGTAA CGTATCGCAA TTCTCTGTAA
751 ACAAAACCAA GGGCTATAAA GGC GTTGCTT TCCCCTTGCC AACAGACGCT
801 GCGGTAGCAA CAGCTACTGG AACAAAGTCT GCGACCATCA ATTATCATGA
851 ATGGCAAGTA GGAGCCTCTC TATCTTACAG ACTAACTCT TTAGTGCCAT
901 ACATTGGAGT ACAATGGTCT CGAGCAACTT TTGATGCTGA TAACATCCGC
951 ATTGCTCAGC CAAAAC TACC TACAGCTGTT TTAAACTTAA CTGCATGGAA
1001 CCCTTCTTTA CTAGGAAATG CCACAGCATT GTCTACTACT GATTCGTTCT
1051 CAGACTTCAT GCAAATTGTT TCCTGTCAGA TCAACAAGTT TAAATCTAGA
1101 AAAGCTTGTG GAGTTACTGT AGGAGCTACT TTAGTTGATG CTGATAAATG
1151 GTCAC TTACT GCAGAAAGCTC GTTTAATTAA CGAGAGAGCT GCTCACGTAT
CTGGTCAGTT CAGATTCTAA

```

The PSORT algorithm predicts an outer membrane location (0.707).

The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 30A) and as a his-tag product. The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 30B) and for FACS analysis (Figure 30C).

45 The cp6998 protein was also identified in the 2D-PAGE experiment (Cpn0695) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6998 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## Example 31

The following *C.pneumoniae* protein (PID 4377102) was expressed <SEQ ID 61; cp7102>:

```

1  MKHTFTKRVL FFFFLVIPIP LLLNLMVVG FFSFAAKANL VQVLHTRATN
51  LSIEFEKKLT IHKLFLDRLA NTLALKSYAS PSAEPYAQAY NEMMALSNNTD
101 FSLCLIDPFD GSVRTKNPGD PFIRYLKQHP EMKKKLSAAV GKAFLLTIPG
151 KPLLHYLILV EDVASWDSTT TSGLLVSFYP MSFLQKDLFQ SLHITKGNIC
201 LVNKYGEVLF CAQDSESSFV FSLDLPNLPQ FQARSPSAIE IEKASGILGG
251 ENLITVSINK KRYLGLVLNK IPIQGTYTLS LVPVSDLIQS ALKVPLNICF
301 FYVLAFLLMW WIFSKINTKL NKPLQELTFC MEAAWRGNHN VRFEPQPYGY
10  351 EFNELGNIFN CTTTTLLNSI EKADIDYHSG EKLQKELGIL SSLQSALLSP
401 DFPTFPKVTF SSQHLRRRQL SGHFNGWTVQ DGGDTLLGII GLAGDIGLPS
451 YLYALSARSL FLAYASSDVS LQKISKDTAD SFSKTTEGNE AVVAMTFIKY
501 VEKDRSLELL SLSEGAPTMF LQRGESFVRL PLETHQALQP GDRLICLTGG
15  551 EDILKYFSQL PIEELLKDPL NPLNTENLID SLTMMLNNET EHSADGTLTI
601 LSFS*

```

A predicted signal peptide is highlighted.

The cp7102 nucleotide sequence <SEQ ID 62> is:

```

1  ATGAACATA CCTTTACCAA GCGTGTCTA TTTTTTTTCT TTTTAGTGAT
51  TCCCATTCCC CTACTCCTCA ATCTTATGGT CGTAGGTTTT TTCTCATTTT
20  101 CTGCCGCTAA AGCAAATTTA GTACAGGTCC TCCATACCCG TGCTACGAAC
151 TTAAGTATAG AATTCGAAAA AAAACTGACG ATACACAAGC TTTTCCTCGA
201 TAGACTTGCC AACACATTAG CCTTAAATC CTATGCATCT CCTTCTGCAG
251 AGCCCTATGC ACAGGCATAC AATGAGATGA TGGCACTCTC CAATACAGAC
301 TTTTCTTAT GCCTTATAGA TCCCTTTGAT GGATCTGTAA GGACGAAAAA
25  351 TCTTGGAGAC CCTTTCATTC GCTATCTAAA ACAGCATCCT GAAATGAAGA
401 AAAAGCTATC CGCAGCTGTA GGGAAAGCCT TTTTATTGAC CATTCCAGGT
451 AAACCACTTT TACATTATCT TATTCTAGTT GAAGATGTCG CATCTTGGGA
501 TTCTACAACG ACTTCAGGAC TGCTTGTAAG TTTCTATCCC ATGTCTTTTT
551 TACAGAAAGA TTTATTCCAA TCCTTACACA TCACCAAAGG AAATATCTGC
30  601 CTGTAAATA AGTATGGCGA GGTCCCTCTC TGTGCTCAGG ACAGTGAATC
651 TTCTTTTGTA TTTTCTCTAG ATCTCCCTAA TTTACCGCAA TTCCAAGCAA
701 GAAGCCCCTC TGCCATAGAA ATTGAGAAAG CTTCTGGAAT TCTTGGTGGG
751 GAGAACCTAA TCACAGTGAG TATCAACAAG AAACGCTACC TAGGATTGGT
801 ACTGAATAAA ATTCCTATCC AAGGGACCTA CACTCTATCT TTAGTTCCAG
35  851 TTTCTGATCT CATCCAATCC GCCTTGAAAG TTCCTCTCAA TATTTGTTTT
901 TTCTATGTAC TTGCTTTTCT CCTCATGTGG TGGATTTTCT CTAAGATCAA
951 CACCAAACCT AACAAAGCCT TTCAAGAACT GACCTTCTGT ATGGAAGCTG
1001 CCTGGCGAGG AAACATAAC GTGAGGTTTG AACCCAGCC TTACGGTTAT
1051 GAATTCAATG AACTAGGAAA TATTTTCAAT TGCACCTCTC TACTCTTATT
40  1101 GAATTCCATT GAGAAAGCAG ATATCGATTA CCATTCAAGC GAAAAATTAC
1151 AAAAAGAATT AGGGATTTTA TCTTCACTAC AAAGTGCCTT ACTAAGTCCG
1201 GATTTCCCTA CGTTCCCTAA AGTTACCTTT AGTTCCCAAC ATCTCCGGAG
1251 AAGGCAACTT TCCGGTCATT TTAATGGTTG GACAGTTCAA GATGGTGGCG
1301 ATACCCTTTT AGGGATCATA GGGCTCGCTG GCGATATTGG TCTTCCTTCC
45  1351 TATCTCTATG CTTTATCCGC ACGGAGTCTT TTTCTTGCCCT ATGCTTCCTC
1401 GGACGTTTCG TTACAAAAAA TCAGCAAGGA TACTGCCGAC AGCTTCTCAA
1451 AAACAACAGA AGGCAATGAG GCTGTAGTTG CTATGACTTT CATTAATAT
1501 GTAGAAAAAG ATCGATCTCT AGAGCTCCTC TCGTTAAGCG AGGGAGCTCC
1551 TACCATGTTT CTACAACGAG GAGAATCTTT CGTACGTCTC CCCTTAGAGA
50  1601 CTCACCAAGC TCTACAGCCT GGAGATCGGT TGATCTGCCT CACTGGAGGA
1651 GAAGACATCC TCAAGTACTT TTCTCAGCTT CCTATTGAAG AGCTCTTAAA
1701 AGATCCTTTA AACCTCTAA ATACAGAGAA TCTTATTGAT TCTCTAACCA
1751 TGATGTTAAA CAACGAAACC GAACATTCTG CAGATGGAAC TCTGACCATC
1801 CTTTCATTTT CATAA

```

55 The PSORT algorithm predicts an inner membrane location (0.338).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product. The purified GST-fusion product is shown in Figure 31A. The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot and for FACS analysis (Figure 31B).

These experiments show that cp7102 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 32

The following *C.pneumoniae* protein (PID 4377106) was expressed <SEQ ID 63; cp7106>:

```

5      1 MKDLGTLGGT SSTAKTVSPD GKVIMGRSQI ADGSWHAFMC HTDFSSNNVL
      51 FDLDNTYKTL RENGRLNSI FNLQNMMLQR ASDHEFTTEFG RSNIALGAGL
     101 YVNALQNLPS NLAAQYFGIA YKIRPKYRLG VFLDHNFFSSH VPNNFNVSHN
     151 RLWMGAFIGW QSDALGSSV KVSFGYGKQK ATITREQLEN TEAGSGESHF
     201 EGVAAQIEGR YGKSLGGHVR VQPFLGLQFV HITRKEYTEN AVQFPVHYDP
10    251 IDYSTGVVYL GIGSHIALVD SLHVGTRMGM EQNFAAHTDR FSGSIASIGN
     301 FVFEKLDVTH TRAFAMRVN YELPYLQSLN LILRVNQPL QGVMGFSSDL
     351 RYALGF*

```

The cp7106 nucleotide sequence <SEQ ID 64> is:

```

15      1 ATGAAAGATT TGGGGACTCT TGGGGGTACC TCTTCTACAG CAAAAACAGT
      51 GTCCCCAGAT GGTAAAGTGA TCATGGGTAG ATCACAAATT GCTGATGGCA
     101 GTTGGCACGC ATTTATGTGT CATACGGATT TCTCCTCTAA TAATGTACTC
     151 TTGATCTCG ATAATACGTA TAAACTCTA AGAGAAAATG GCCGTCAGCT
     201 AAATTCCATA TTCAACCTAC AAAATATGAT GTTACAGAGA GCCTCAGATC
     251 ATGAGTTCAC AGAGTTTGA AGGAGTAACA TCGCTCTTGG TGCCGGGCTT
20    301 TATGTGAATG CCTGTCAGAA TCTCCCTAGC AATTTAGCAG CACAATATTT
     351 TGGAATCGCA TACAAAATAC GTCCTAAATA TCGTTTGGGG GTGTTTTTGG
     401 ACCATAATTT CAGCTCCAC GTTCCTAATA ATTTTAACGT AAGCCACAAT
     451 AGACTCTGGA TGGGAGCCTT TATTGGATGG CAGGATTCTG ATGCTCTAGG
     501 ATCTAGTGTC AAGGTGTCTT TCGGATATGG AAAACAAAAA GCCACGATTA
25    551 CAAGAGAGCA ATTAGAGAAT ACAGAAGCCG GGAGTGGGGA GAGCCATTTT
     601 GAAGGGGTCG CTGCTCAGAT AGAAGGGCGG TATGGTAAGA GCCTCGGAGG
     651 ACATGTCAGG GTCCAGCCTT TCCTAGGACT GCAGTTTGTC CACATTACAA
     701 GGAAAGAATA TACCGAAAAT GCAGTGCAAT TTCCTGTACA CTATGATCCT
     751 ATAGACTATT CTACAGGTGT AGTGTATTTA GGAATTGGAT CTCATATTGC
30    801 ACTTGTAGAT TCTTTACATG TAGGCACACG CATGGGAATG GAGCAAAACT
     851 TTGCAGCCCA TACGGACAGG TTCTCAGGAT CTATAGCGTC TATTGGAAAC
     901 TTTGTGTTTG AAAAGCTTGA TGTGACTCAC ACAAGGGCAT TTGCGGAAAT
     951 GCGTGTCAAC TATGAGCTTC CCTATCTACA GTCTCTGAAT CTTATCTTAC
35   1001 GAGTTAATCA ACAGCCTCTA CAAGGGGTTA TGGGATTTTC CAGTGATCTT
     1051 AGGTATGCCT TAGGATTCTA A

```

The PSORT algorithm predicts a cytoplasmic location (0.224).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product. The purified GST-fusion product is shown in Figure 32A. The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 32B) and for FACS analysis (Figure 32C).

This protein also showed very good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7106 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 33

The following *C.pneumoniae* protein (PID 4377228) was expressed <SEQ ID 65; cp7228>:

```

      1 MTAVLILTSF PSEESARSLA RHLITERLAS CVHVFPKGTS TYLWEGKLCE
     51 SEEHHIQIKS IDIRFSEICL AIQEFSGYEV PEVLLFFPIEN GDPRLYNLWT
    101 ILSYPEKPPL SD*

```

The cp7228 nucleotide sequence <SEQ ID 66> is:

```

1   ATGACTGCTG TTCTTATTCT TACATCTTTC CCTTCGGAGG AAAGTGCTCG
51  CTCCTTAGCT AGACATCTGA TTACAGAGCG TCTTGCTTCC TGTGTGCATG
101 TATTCCCTAA AGGCACATCG ACATATCTAT GGAAGGCAA GCTATGTGAG
5   151 TCTGAAGAAC ATCATATACA AATCAAATCG ATAGACATAC GCTTCTCGGA
201 AATTTGTCTT GCTATTCAGG AGTTCTCTGG CTATGAGGTT CCTGAAGTCT
251 TACTATTTCC TATTGAAAAT GGGATCCGA GGTACTTGAA TTGGTTAACG
301 ATTCTCAGCT ATCCAGAGAA GCCTCCGCTT TCAGATTAG

```

The PSORT algorithm predicts an inner membrane location (0.040).

- 10 The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 33A (his-tag = left-hand arrow, GST = right-hand arrow). The proteins were used to immunise mice, whose sera were used in a Western blot (Figure 33B) and FACS analysis.

These experiments show that cp7228 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### 15 Example 34

The following *C.pneumoniae* protein (PID 4377170) was expressed <SEQ ID 67; cp7170>:

```

1   MNSKMLKHLR LATLSFSMFF GIVSSPAVYA LGAGNPAAPV LPGVNPEQTG
51  WCAFQLCNSY DLFAALAGSL KFGFYGDYVF SESAHITNVP VITSVTTSQT
20  101 GTTPTITSTT KNVDFDLNNS SSSSCVFAT IALQETSPAA IPLLDIAFTA
151 RVGGLKQYYR LPLNAYRDFT SNPLNAESEV TDGLIEVQSD YGIVWGLSLQ
201 KVLWKDGVSF VGVSA DYRHG SSPINYIIVY NKANPEIYFD ATDGNLSYKE
251 WSASIGISTY LNDYVLPYAS VSIGNTSRKA PSDSFTELEK QFTNFKFKIR
301 KITNFDRVNF CFGTTCCISN NFYYSVEGRW GYQRAINITS GLQF*

```

A predicted signal peptide is highlighted.

- 25 The cp7170 nucleotide sequence <SEQ ID 68> is:

```

1   ATGAATAGCA AGATGCTAAA ACATTTACGT TTAGCAACCC TTTCCTTCTC
51  TATGTTCTTC GGGATTGTAT CTTCTCCCGC AGTATATGCC CTAGGGGCTG
101 GAAACCCTGC AGCTCCAGTA CTCCCAGGTG TGAATCCTGA GCAAACGGGA
30  151 TGGTGTGCCT TCCAACCTTG TAATAGTTAC GATCTTTTGT CTGCTCTTGC
201 AGGAAGCCTC AAATTTGGGT TCTATGGAGA TTATGTCTTC TCAGAAAGTG
251 CCCATATTAC CAATGTCCCT GTCATTACCT CCGTTACGAC TTCAGGCACA
301 GGAACAACGC CAACCAATTAC CTCTACAAC TAAAAACGTAG ACTTTGATCT
35  351 TAACAACAGC TCCATCAGCT CGAGCTGTGT TTTTGCAACC ATAGCTCTAC
401 AGGAAACATC CCCAGCTGCC ATTCCCCTTT TAGATATAGC CTTCACTGCA
45  451 CGTGTGCGAG GACTTAAGCA GTACTACCGC CTCCCTCTCA ATGCTTACAG
501 AGACTTCACT TCAAATCCTT TAAATGCAGA ATCTGAAGTT ACAGATGGTC
551 TCATTGAAGT CCAGTCAGAC TATGGAATG TCTGGGGTCT GAGTTTACAA
601 AAAGTATTGT GGAAAGATGG AGTGTCTTTT GTAGGGGTGA GCGGTGACTA
651 CCGTCACGGT TCCAGTCCCA TCAACTATAT CATCGTTTAC AACAAGGCCA
40  701 ACCCCGAGAT CTATTTTCGAT GCTACTGATG GAAACCTAAG CTATAAAGAA
751 TGGTCTGCAA GCATCGGCAT CTCTACGTAT CTTAATGACT ATGTGCTTCC
801 CTATGCATCC GTATCTATAG GAAATACTTC AAGAAAAGCT CCTTCTGATA
851 GCTTCACAGA ACTCGAAAAG CAATTTACGA ATTTTAAATT TAAAATTCGT
901 AAAATCACAA ACTTCGACAG AGTAAACTTC TGCTTCGGAA CTACCTGCTG
45  951 CATCTCAAAT AACTTCTACT ATAGTGTAGA AGGCCGTTGG GGATATCAGC
1001 GTGCTATCAA CATTACGTCA GGTCTGCAGT TTTAG

```

The PSORT algorithm predicts a bacterial outer membrane location (0.936).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product. The purified GST-fusion product is shown in Figure 34A. The GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (34B) and for FACS analysis (34C).



The cp7170 protein was also identified in the 2D-PAGE experiment (Cpn0854).

These experiments show that cp7170 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 35

5 The following *C.pneumoniae* protein (PID 4377072) was expressed <SEQ ID 69; cp7072>:

```

1  MDIKLFLCLF LCSSLIAMSE IYGKTGDYEK LTLTGINIID RNGLSETICS
51  KEKLLKKYTKV DFLAPQPYQK VMRMVKNKRG DNVSCLTAYH TNGQIKQYLE
101 CLNNRAYGRY REWHVNGNIK IQAEVIGGIA DLHPSAESGW LFDQTTTFAYN
151 DEGILEAAIV YEKGLLEGSS VYYHTNGNIW KECPYHKQVP QGKFLTYTSS
10  201 GKLLKEQNYQ QGKRHGLSIR YSEDSEEDVL AWEEYHEGRL LKAEYLDPQT
251 HEIYATIHEG NGIQAIYGY AVIETRAFYP GEPLYGVTRF DNSGTQIVQT
301 YNLLQGAKHG EEEFFYPETG KPKLLLNWHE GILNGIVKTW YPGGTLESCK
351 ELVNNKKSGL LTIYYPEGQI MATEEYDNDL LIKGEYFRPG DRHPYSKIDR
401 GCGTAVFFSS AGTITKKIPY QDGKPLLN*

```

15 A predicted signal peptide is highlighted.

The cp7072 nucleotide sequence <SEQ ID 70> is:

```

1  ATGGATATAA AAAAAGTCTT TTGCTTATTT CTATGTTCTT CTCTAATTGC
51  CATGAGTCCC ATTTATGGGA AAACAGGTGA CTATGAGAAA CTCACCCCTTA
101 CAGGGATCAA TATCATTGAT AGAAACGGCC TGTCAGAAAC TATTTGCTCT
20  151 AAAGAGAAGC TAAAGAAATA CACCAAGGTA GACTTTCTTG CTCCCCAGCC
201 CTATCAAAAG GTCATGAGGA TGTATAAAAA CAAACGCGGA GATAACGTTT
251 CTTGTTTAAAC AGCCTATCAC ACTAACGGGC AAATTAAGCA GTACCTGGAG
301 TGTCTCAATA ATCGTGCTTA TGGAAGATAT CGTGAATGGC ACGTCAACGG
25  351 GAATATCAAA ATCCAAGCTG AGGTTATCGG AGGTATTGCG GATCTTCATC
401 CCTCAGCAGA GTCTGGCTGG CTATTTGATC AAACACATT TGCTTATAAT
451 GATGAAGGTA TCTTAGAAGC CGCTATCGTC TATGAAAAAG GGCTGCTCGA
501 AGGATCTTCG GTGTATTACC ATACTAATGG GAATATTGG AAAGAGTGTG
551 CCTATCATAA GGGAGTTCCCT CAAGGTAAAT TCCTGACATA CACATCTTCG
601 GGGAAACTGC TCAAAGAACA GAATTACCAA CAAGGCAAAA GACACGGTCT
30  651 TTCGATTTCG TACAGCGAAG ATTCCGAAGA AGATGTTTAA GCCTGGGAAG
701 AATATCATGA GGGACGACTC CTAAAAGCAG AGTACTTAGA TCCTCAAAC
751 CACGAAATCT ATGCGACTAT ACACGAAGGG AACGGCATTC AAGCAATCTA
801 CGGCAAGTAT GCCGTTATAG AAAC TAGGGC ATTTTACCGA GGGGAACCTT
851 ATGGAAGTAT TACCAGATTC GACAACCTCCG GAACACAGAT TGTCCAAACG
35  901 TATAACCTTT TGCAAGGCGC GAAGCACGGA GAAGAATTTT TCTTTTATCC
951 TGAGACAGGG AAACCCAAGC TGCTTCTTAA TTGGCATGAA GGAATTTTAA
1001 ATGGGATAGT AAAAAGTCTG TATCCCGGAG GAACCTTAGA AAGTTGTAAA
1051 GAACTCGTAA ATAACAAAAA ATCCGGGTTA CTGACCATT ACTACCCTGA
1101 AGGACAGATC ATGGCGACCG AAGAGTATGA TAATGATCTT CTAATTAAAG
40  1151 GAGAGTACTT CCGCCCTGGA GACCGTCATC CCTACTCTAA AATAGATCGT
1201 GATTGTGGGA CTGCAGTAT TTTCTCGTCG GCGGGAAC TA TACTAAAAA
1251 AATCCCCAT CAGGACGGCA AACCTTTGCT CAACTAG

```

The PSORT algorithm predicts a periplasmic location (0.688).

45 The protein was expressed in *E.coli* and purified as a his-tag product (Figure 35A) and as a GST-fusion product (Figure 35B). The recombinant his-tag protein was used to immunise mice, whose sera were used in a Western blot (Figure 35C) and for FACS analysis.

These experiments show that cp7072 is a useful immunogen. These properties are not evident from the sequence alone.

### Example 36

50 The following *C.pneumoniae* protein (PID 4376879) was expressed <SEQ ID 71; cp6879>:

-76-

1 MATPAQKSPT FQDPSFVREL GSNHPVFSPL TLEERGEMAI ARVQQCGWNH  
 51 TIVKVSILIL ALLTILGGGL LVGLLPVPM FIGTGLIALG AVIFALALIL  
 101 CLYDSQGLPE ELPPVPEPQQ IQIEDLRNET REVLEGLLE VLLKDRDAKD  
 151 PAVPQVVVDC EKRLGMLDRK LRREEEILYR STAHLKDEER YEFLLELLEM  
 5 201 RSLVADRLEF NRRSYERFVQ GIMTVRSEEG EKEISRLQDL ISLQQQTQVD  
 251 LRSRIDDEQK RCWTALQRIN QSQKDIQRAH DREASQRACE GTEMDCAEERQ  
 301 QLEKDLRRQL KSMQEWIEMR GTIHQQEKAW RKQNAKLERL QEDLRLTGIA  
 351 FDEQSLFYRE YKEKYSQKL DMQKILQEVN AEKSEKACLE SLVHDYEQQL  
 401 EQKDANLKA AAVWEEELGK QQQEDYEQTQ EIRRLSTFIL EYQDSLREAE  
 10 451 KVEKDFQELQ QRYRSLQEEK QVKEKILEES MNHFADLF EK AQKENMAYKK  
 501 KLADLEGAAA PTEIGEDDDW VLTDSASLSQ KKIRELVEEN QELLKALAFK  
 551 SNELTQLVAD AVEAEKEISK LREHIEEQKE GLRALDKMHA QAIKDCEAAQ  
 601 RKCCDLESLL SPVREDAGMR FELEVELQRL QEENALRAE VERLEQEQQFQ  
 651 G\*

15 The cp6879 nucleotide sequence <SEQ ID 72> is:

1 ATGGCAACAC CCGCTCAAAA ATCCCTTACA TTTCAAGATC CTAGTTTTGT  
 51 AAGAGAGCTA GGCAGTAACC ACCCTGTCTT TTCCCGCTA ACCTTGAGG  
 101 AAAGAGGGGA GATGGCAATA GCTCGAGTCC AGCAGTGTGG ATGGAATCAT  
 151 ACAATTGTTA AGGTAAGTCT TATTATTCTT GCTCTTCTTA CTATTTTAGG  
 20 201 GGGAGGATTA CTCGTAGGAT TGCTGCCAGC AGTTCCTATG TTTATTGGAA  
 251 CAGGTCTGAT TGCTTTGGGA GCCGTATATAT TTGCTTTGGC TTTGATTTTA  
 301 TGTCTTTATG ATTCTCAGGG CCTTCCTGAG GAACTCCCTC CGGTTCTGA  
 351 ACCACAACAA ATTCTAGATTG AAGATTTAAG AAACGAGACC AGAGAAGTTC  
 401 TTGAAGGGAC TCTTTTAGAG GTTCTCTTAA AGGATAGAGA CGCTAAGGAC  
 25 451 CCTGCGGTGC CCCAGGTGGT TGTAGACTGT GAAAAGCGTC TTGGAATGTT  
 501 GGATCGTAAG CTGCGACGTG AAGAGGAGAT TCTGTATCGC TCGACGGCCC  
 551 ATCTTAAAGA CGAGGAAAGG TATGAGTTCT TGCTGGAGCT CTTGGAAATG  
 601 CGTAGTCTGG TTGCCGATCG GCTAGAATTT AACCCTAGAA GTTATGAGCG  
 651 ATTTGTTCAA GGAATTATGA CAGTTAGATC AGAGGAGGGG GAAAAAGAGA  
 30 701 TTTCTCGTCT ACAAGATCTA ATCAGTTTGC AGCAGCAGAC GGTGCAAGAT  
 751 TTAAGGAGTC GGATCGATGA CGAGCAGAAG AGATGCTGGA CGGCTTTACA  
 801 ACGTATTAAC CAATCTCAGA AGGATATACA ACGGGCTCAT GATCGCGAGG  
 851 CTTCGCAGCG TGCTGTGAG GGCACAGAGA TGGATTGTGC AGAACGCCAG  
 901 CAACTGGAGA AGGATTTAAG GAGACAGCTG AAATCTATGC AGGAGTGGAT  
 35 951 TGAGATGAGG GGCACAATCC ATCAACAAGA GAAGGCTTGG CGTAAGCAGA  
 1001 ATGCCAAATT AGAAAGATTA CAAGAGGATC TGAGACTTAC TGGGATTGCT  
 1051 TTTGACGAAC AATCTCTGTT CTATCGCGAA TATAAAGAGA AATATCTGAG  
 1101 TCAGAAACTA GATATGCAAA AGATTTTACA GGAAGTCAAC GCAGAGAAAA  
 1151 GTGAGAAGGC TTGCTTAGAG AGTCTGGTCC ATGACTATGA GAAGCAGCTC  
 40 1201 GAACAAAAAG ATGCTAATCT GAAGAAAGCA GCAGCTGTTT GGAAGAAGA  
 1251 ATTAGGGAAG CAGCAACAGG AAGACTACGA ACAAACCCAA GAAATTAGAC  
 1301 GTCTGAGTAC ATTCATTCTT GAGTACCAG ACAGTCTGCG TGAGGCAGAA  
 1351 AAAGTTGAGA AAGATTTCCA AGAGCTACAA CAAAGGTATA GCCGTCTTCA  
 1401 AGAGGAGAAA CAGGTAAAAG AAAAAATCTT AGAAGAAAGT ATGAATCATT  
 45 1451 TTGCCGATCT CTTTGAGAAG GCTCAAAAGG AAAACATGGC CTACAAGAAG  
 1501 AAGTTAGCGG ATTTAGAGGG TGCCGCTGCT CTTACTGAGA TCGGTGAGGA  
 1551 CGATGACTGG GTACTCACAG ATTCTGCTTC TCTCAGCCAG AAGAAGATCC  
 1601 GCGAACTCGT GGAAGAGAAT CAAGAATCC TGAAAGCACT TGCATTTAAA  
 1651 TCTAACGAAT TGACTCAACT GGTGCGCGAT GCTGTAGAAG CTGAAAAAGA  
 50 1701 AATCAGCAAG CTTCGAGAAC ACATAGAAGA GCAGAAAGAA GGATTACGAG  
 1751 CTCTTGATAA GATGCATGCA CAAGCGATCA AAGATTGCGA AGCTGCTCAG  
 1801 AGAAAAATGCT GTGACCTTGA GAGCCTTCTC TCTCCTGTTC GAGAAGATGC  
 1851 TGAATGAGA TTTGAGCTAG AGGTCGAGCT TCAAAGATTG CAAGAAGAAA  
 1901 ATGCACAGCT TAGAGCGGAG GTTGAAAGAC TAGAGCAAGA GCAATTTCAA  
 55 1951 GGATAA

The PSORT algorithm predicts an inner membrane location (0.646).

The protein was expressed in *E. coli* and purified as a his-tag product and as a GST-fusion product. The purified GST-fusion product is shown in Figure 36A. The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 36B) and for FACS analysis.

60 These experiments show that cp6879 is useful immunogen. These properties are not evident from the sequence alone.

**Example 37**

The following *C.pneumoniae* protein (PID 4376767) was expressed <SEQ ID 73; cp6767>:

```

1  MIKQIGRFFR AFIFIMPLSL TSCESKIDRN RIWIVGTNAT YPPFEYVDAQ
51  GEVVGFDIDL AKAISEKLGK QLEVREFAFD ALILNLKKHR IDAILAGMSI
101 TPSRQKEIAL LPYYGDEVQE LMVVSRSLE TPVPLPTQYS SVAVQTGTFQ
151 EHYLLSQPGI CVRSFDSTLE VIMEVRYGKS PVAVLEPSVG RVVLKDFPNL
201 VATRLELPPE CWVLGCGLGV AKDRPEEIQT IQQAITDLKS EGVIIQSLTKK
251 WQLSEVAYE*
```

The cp6767 nucleotide sequence <SEQ ID 74> is:

```

10  1  ATGATAAAAC AAATAGGCCG TTTT TTTTAGA GCATTTATTT TTATAATGCC
51  TTTATCTTTA ACAAGTTGTG AGTCTAAAAT CGATCGAAAT CGCATCTGGA
101 TTGTAGGTAC GAATGCTACA TATCCTCCTT TTGAGTATGT GGATGCTCAG
151 GGGGAAGTTG TAGGTTTCGA TATAGATTG GCAAAGGCAA TTAGTGAAAA
201 ACTTGGCAAG CAATTGGAAG TTAGAGAATT CGCTTTCGAT GCTTTAATTT
15  251 TAAATTTAAA AAAACATCGT ATCGATGCAA TTTTAGCAGG AATGTCCTAT
301 ACTCCTTCGC GTCAGAAGGA AATCGCCCTG CTTCCCTATT ATGGCGATGA
351 GGTTCAGAG CTGATGGTGG TTTCTAAGCG GTCTTTAGAG ACCCCTGTGC
401 TTCCCTAAC ACAGTATTCT TCTGTTGCTG TTCAGACAGG AACGTTTCAG
451 GAGCATTATC TTTTATCTCA GCCCGGAATT TGTGTCCGTT CTTTGTAGAT
20  501 CACCTTGGAG GTGATTATGG AAGTTCGTTA TGGGAAATCT CCGGTTGCCG
551 TTCTAGAACC CTCGGTAGGA CGTGTCTGTC TTAAAGACTT CCCTAATCTT
601 GTTGCAACAA GATTAGAGCT CCCTCCTGAA TGTGTGGTGT TGGGCTGTGG
651 TCTCGGCGTA GCTAAAGATC GTCCTGAAGA AATACAAACG ATTCAACAAG
701 CGATTACAGA TTAAAGAGC GAAGGGGTGA TTCAATCTTT AACCAAGAAA
25  751 TGGCAACTTT CTGAAGTTGC TTACGAATAG
```

The PSORT algorithm predicts an inner membrane location (0.083).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product. The purified his-tag product is shown in Figure 37A. The recombinant his-tag protein was used to immunise mice, whose sera were used in a Western blot (Figure 37B) and for FACS analysis (Figure 37C). The GST-fusion was also used in a Western blot (Figure 37D).

The cp6767 protein was also identified in the 2D-PAGE experiment and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6767 is a useful immunogen. These properties are not evident from the sequence alone.

**Example 38**

The following *C.pneumoniae* protein (PID 4376717) was expressed <SEQ ID 75; cp6717>:

```

1  MMSRLRFRLA ALGIFFILLV PNSVSAKTIV ASDKEKVGVL VYDNSVEAFQ
51  QILDCIDHAN FYVELCPCMT GGRTLKEMVD HLEARM DLVP ELCSYIIIQP
101 TFTDAEDQKL LKALKERHPN RFFYVFTGCP PSTSILAPNV IEMHIKLSII
40  151 DGKYCILGGT NFEEFMCPTG DEVPEKVDNP RLFVSGVRRP LAFRDQDIML
201 RSTAFGLQLR BEYHKQFAMW DYYAHMWF I DNPEQFAGAC PPLTLEQAE
251 TVFPGFDKHE DLVLVDSSKI RIVLGGPHDK QPNPVTQEYL KLIQGARSSV
301 KLAHMYFIPK DELLNALVDV SHNHGVHLSL ITNGCHELSP AITGPYAWGN
351 RINYFALLYG KRYPLWKKWF CEKLKPYERV SIYEFAIWET QLHKKCMIID
45  401 DEIFVIGSYN FGKKSADFY ESIVVIESPE VAAKANKVFN KDIGLSIPVS
451 HGDIFSWYFH SVHHTLGH LQ LTYMPA*
```

A predicted signal peptide is highlighted.

The cp6717 nucleotide sequence <SEQ ID 76> is:

1 ATGATGAGTC GGTGCGTTT TCGCTTGGCA GCTCTTGGAA TATTTTTTAT  
 51 TTTGCTGGTT CCTAATTCTG TTTTCAGCAA GACAATCGTA GCTTCAGACA  
 101 AGGAGAAGGT TGGAGTTCTT GTTTATGACA ATAGTGTAGA GGCCTTTCAA  
 151 CAGATATTGG ATTGCATAGA TCATGCAAAAT TTTTATGTAG AACTGTGTCC  
 5 201 CTGCAATGACA GGAGGCCGAA CGCTTAAAGA GATGGTAGAT CACCTCGAGG  
 251 CTCGTATGGA TCTGGTTCCA GAGCTCTGTA GCTATATCAT TATCCAACCC  
 301 ACGTTTACCG ATGCTGAAGA CCAAAAATTA CTCAAAGCTC TCAAAGAACG  
 351 TCATCCCAAC CGGTTTTTCT ACGTTTTTAC AGGGTGCCCA CCCTCAACAA  
 401 GCATCCTCGC TCCTAATGTC ATTGAAATGC ATATCAAAC TTTCTATCATC  
 10 451 GATGGGAAAT ATTGTATTTT AGGTGGTACC AATTTTGAAG AGTTTATGTG  
 501 CACTCCAGGG GATGAGGTTT CTGAGAAAGT GGATAACCCA CGTTTATTTG  
 551 TCAGTGGAGT GCGTCGCCCC CTAGCATTTT GTGATCAGGA TATCATGTTG  
 601 CGTTCTACAG CATTCGGTTT GCAGCTCAGA GAAGAATATC ATAAGCAATT  
 651 TGCTATGTGG GACTACTATG CACATCATAT GTGGTTCATT GATAATCCTG  
 15 701 AACAGTTTGC AGGCGCCTGT CCTCCACTGA CTTTAGAACA AGCCGAGGAG  
 751 ACAGTATTTT CTGGATTGTA CAAACATGAA GATCTTGTTT TTGTCGACTC  
 801 TTCCAAGATC AGGATAGTTT TAGGTGGTCC CCACGATAAG CAACCCAATC  
 851 CTGTGACTCA AGAATATTTG AAACCTATCC AGGGAGCTAG ATCTTCTGTG  
 901 AAGCTTGCTC ACATGTATTT CATCCCTAAG GACGAGCTTT TAAATGCTCT  
 20 951 TGTCGACGTT TCTCATAATC ACGGTGTTCA TCTGAGTTTA ATTACGAACG  
 1001 GCTGTCATGA ATTAAGTCCT GCAATTACAG GACCCTATGC TTGGGGAAAC  
 1051 CGTATTAACT ATTTCGCCTT GCTCTATGGG AAACGGTATC CTCTTTGGAA  
 1101 AAAATGGTTT TGCGAAAAGC TAAAACCTTA TGAGCGGGTT TCTATTTATG  
 1151 AGTTTGCTAT TTGGGAAACG CAGTTGCACA AGAAGTGTAT GATTATCGAT  
 25 1201 GATGAAATTT TTGTGATCGG AAGTTATAAT TTTGGAAAGA AAAGTGATGC  
 1251 CTTTGATTAC GAAAGTATTG TAGTTATCGA ATCTCCAGAA GTCGCTGCAA  
 1301 AAGCTAACAA AGTCTTCAAT AAAGATATCG GATTGTGAT TCCTGTAAGT  
 1351 CATGGCGACA TTTTCTCTTG GTATTTCAT TCCGTACACC ACACTTTGGG  
 1401 ACATTTGCAG CTGACCTATA TGCCAGCCTA G

30 The PSORT algorithm predicts a periplasmic location (0.939).

The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 38A), as a his-tagged protein, and as a GST/his fusion product. The proteins were used to immunise mice, whose sera were used in a Western blot (Figure 38B) and for FACS analysis.

35 These experiments show that cp6717 is a useful immunogen. These properties are not evident from the sequence alone.

### Example 39

The following *C.pneumoniae* protein (PID 4376577) was expressed <SEQ ID 77; cp6577>:

1 MKKLLFSTFL LVLGSTSAAH ANLGYVNLKR CLEESDLGKK ETEELEAMKQ  
 51 QFVKNAEKIE BELTSIYNKL QDEDYMESLS DSASEELRKK FEDLSGEYNA  
 40 101 YQSQQYQSIN QSNVKRIQKL IQEVKIAAES VRSKEKLEAI LNEEAFLAIA  
 151 PGTDKTTEII AILNESFKKQ N\*

A predicted signal peptide is highlighted.

The cp6577 nucleotide sequence <SEQ ID 78> is:

1 ATGAAAAAAT TATTATTTTC TACATTTCTT CTTGTTTTAG GATCAACAAG  
 45 51 CGCAGCTCAT GCAAATTTAG GCTATGTTAA TTTAAAGCGA TGTCTTGAAG  
 101 AATCCGATCT AGGTAAAAAG GAAACTGAAG AATTGGAAGC TATGAAACAG  
 151 CAGTTTGTAA AAAATGCTGA GAAAATAGAA GAAGAACTCA CTTCTATTTA  
 201 TAATAAGTTG CAAGATGAAG ATTACATGGA AAGCCTATCG GATTCTGCCT  
 251 CTGAAGAGTT GCGAAAGAAA TTGGAAGATC TTTTCAGGAG GTACAATGCC  
 50 301 TACCAGTCTC AGTACTATCA ATCTATCAAT CAAAGTAATG TAAAACGCAT  
 351 TCAAAAACCTC ATTCAAGAAG TAAAAATAGC TGCAGAAATCA GTGCGGTCCA  
 401 AAGAAAAACTC AGAAGCTATC CTTAATGAAG AAGCTGTCTT AGCAATAGCA  
 451 CCTGGGACTG ATAAAACAAC CGAAATTATT GCTATTCTTA ACGAATCTTT  
 501 CAAAAACAA AACTAG

55 The PSORT algorithm predicts a periplasmic space location (0.932).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 39A) and as a GST-fusion product (Figure 39B). The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 39C) and for FACS analysis.

The cp6577 protein was also identified in the 2D-PAGE experiment.

- 5 These experiments show that cp6577 is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 40

The following *C.pneumoniae* protein (PID 4376446) was expressed <SEQ ID 79; cp6446>:

```

10      1  MKQPMSLIFS SVCLGLGLGS LSSCNQKPSW NYHNTSTSEE FFFVHGKNSVS
      51  QLPHPYSAFR TTQIFSEEHN DPYVVAKTDE ESRKIWREIH KNLKIKGSYI
     101  PISTYGSIMH PKSAAITLKT YRPHPIWING YERSFNIDTG KYLKNRSRRR
     151  TSHDGPKNRA VLNLKSSGR RCNAIGLEMT EEDFVIARRR EGVVSLYPVE
     201  VCSYPQGNPF VIAYAWIADE SACSKEVLPV KGYVSLVWES VSSDSLNAF
     251  GDSFAEDYLR STFLANGTSI LCVHESYKKV PPQP*
```

- 15 A predicted signal peptide is highlighted.

The cp6446 nucleotide sequence <SEQ ID 80> is:

```

      1  ATGAAACAGC CCATGTCTCT TATCTTTTCA AGTGTATGTT TAGGATTAGG
      51  TCTTGGATCT CTTTCCTCCT GTAATCAAAA GCCCTCTTGG AATTATCACA
     101  ACACITCAAC GAGCGAAGAA TTCITTGTTC ATGGAAATAA GAGTGTTTCG
     201  CAACITGCTC ATTATCCTTC TGCATTTCGT ACGACTCAAA TCTTTTCTGA
     251  AGAGACAAAT GATCCTTATG TCGTAGCTAA GACTGATGAA GAGTCTCGTA
     301  AAATTTGGAG AGAAATCCAT AAAAATCTCA AAATCAAAGG TTCTTACATT
     351  CCCATATCGA CTTATGGAAG TCTGATGCAC CCAAAATCAG CAGCTCTTAC
     401  ATTAAAAACG TATCGTCCAC ATCCTATTTG GATAAATGGA TACGAGCGTT
     451  CTTTTAATAT AGACACAGGA AAGTACTTAA AAAACGGAAG TCGCCGTAGA
     501  ACTTCTCACG ATGGTCCGAA AAATCGAGCT GTACTGAATC TCATTAAATC
     551  TTCGGGACGA CGCTGTAATG CTATAGGCCT TGAGATGACA GAAGAAGACT
     601  TTGTAATAGC TAGAAGGCGA GAAGGTGTTT ATAGCCTGTA TCCCGTTGAA
     651  GTGTGCTCGT ATCCTCAGGG GAATCCTTTT GTCATTGCTT ATGCCCTGGAT
     701  TGCAGATGAG AGTGCCTTGT CAAAAGAGGT CCTACCTGTA AAAGGGTACT
     751  ATTCTTTAGT CTGGGAAAGC GTTCTCTCCT CTGATTCTCT GAATGCTTTT
     801  GGAGATTCCT TTGCAGAGGA CTACCTCAGA AGCACGTTT TAGCAAACGG
     851  AACTTCTATA CTCTGTGTTC ATGAAAGCTA TAAGAAAGTT CCTCCTCAGC
     901  CCTAA
```

- 35 The PSORT algorithm predicts an inner membrane location (0.177).

The protein was expressed in *E.coli* and purified as a his-tag product and a GST-fusion product. The GST-fusion product is shown in Figure 40A. The recombinant his-tag protein was used to immunise mice, whose sera were used in a Western blot (Figure 40B) and for FACS analysis.

- 40 These experiments show that cp6446 is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 41

The following *C.pneumoniae* protein (PID 4377108) was expressed <SEQ ID 81; cp7108>:

```

45      1  MSKKIKVLGH LTLCTLFRGV LCAAALSNIG YASTSQESPY QKSIEDWKGY
      51  TFTDLELLSK EGWSEAHAVS GNGSRIVGAS GAGQGSVTAV IWESHLIKHL
     101  GTLGGEASSA EGISKDGEVV VGWSDTREGY THAFVFDGRD MKDLGTLGAT
     151  YSVARGVSGD GSIIVGVSAT ARGEDYGWQV GVKWEGKIK QLKLLPQGLW
```

-80-

201 SEANAISEDG TVIVGRGEIS RNHIVAVKWN KNAVYSLGTL GGSVASAEAI  
 251 SANGKVIVGW STTNNGETHA FMHKDETMHD LGTLGGGFSV ATGVSADGRA  
 301 IVGFSAVKTG EIHAFYYAEG EMEDLTTLGG BEARVFDISS EGNDIIGSIK  
 351 TDAGAERAYL FHIHK\*

5 A predicted signal peptide is highlighted.

The cp7108 nucleotide sequence <SEQ ID 82> is:

1 ATGAGTAAGA AGATAAAGGT TCTAGGTCAT TTGACGCTCT GCACTCTGTT  
 51 TAGAGGAGTG CTGTGTGCAG CGGCCCTTTC CAACATAGGA TATGCGAGTA  
 101 CTTCTCAGGA ATCACCATAT CAGAAGTCTA TAGAAGACTG GAAAGGGTAT  
 151 ACCTTTACAG ATCTTGAGTT ACTGAGTAAG GAAGGGTGGT CTGAAGCTCA  
 201 TGCAGTTTCT GGAAATGGCA GTAGAATTGT AGGAGCTTCG GGAGCTGGCC  
 251 AAGGTAGTGT GACTGCTGTC ATATGGGAAA GTCACCTGAT AAAACATCTC  
 301 GGCACCTTAG GTGGCGAGGC TTCATCTGCA GAGGGAATTT CAAAGGATGG  
 351 AGAGGTGGTC GTTGGGTGGT CAGATACTAG AGAGGGATAT ACTCATGCCT  
 401 TTGTCTTCGA CGGTAGAGAT ATGAAAGATC TCGGTACTCT AGGAGCTACC  
 451 TATTCTGTAG CAAGGGTGT TTCTGGAGAT GGTAGTATCA TCGTAGGAGT  
 501 CTCTGCAACT GCTCTGGAG AGGATTACGG ATGGCAAGTT GGTGTCAAGT  
 551 GGGAAAAAGG GAAAATCAAA CAATTGAAGT TGTTCCTCA AGGTCTCTGG  
 601 TCTGAGGCGA ATGCAATCTC TGAGGATGGT ACGGTGATTG TCGGGAGAGG  
 651 GGAAATCTCT CGCAATCACA TCGTTGCTGT AAAATGGAAT AAAAATGCTG  
 701 TGTATAGTTT GGGGACTCTC GGAGGTAGTG TCGCTTCAGC AGAGGCTATA  
 751 TCGGCAAATG GGAAAGTAAT TGAGGATGG TCCACGACTA ATAATGGTGA  
 801 GACTCATGCC TTTATGCACA AAGATGAGAC AATGCACGAT CTCGGCACTC  
 851 TAGGAGGAGG TTTTCTGTC GCAACTGGAG TTTCTGCTGA TGGGAGAGCC  
 901 ATCGTAGGAT TTTCAGCAGT GAAGACCGGA GAAATTCATG CTTTTTACTA  
 951 TGCAGAAGGA GAAATGGAGG ATTTAACAAC TTTGGGAGGG GAAGAAGCTC  
 1001 GAGTGTTCTGA CATATCTAGC GAAGGAAACG ATATCATTTG CTCTATAAAA  
 1051 ACTGACGCTG GAGCTGAACG CGCCTATCTG TTCCATATAC ATAAATAA

The PSORT algorithm predicts an outer membrane location (0.921).

30 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 41A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 41B) and for FACS analysis (Figure 41C). A his-tagged protein was also expressed.

The cp7108 protein was also identified in the 2D-PAGE experiment.

These experiments show that cp7108 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 42

The following *C.pneumoniae* protein (PID 4377287) was expressed <SEQ ID 83; cp7287>:

1 MVAKKTVRSY RSSFHSVIV AILSAGIAFE AHSLHSSELD LGVFNKQFEE  
 51 HSAHVVEAQT SVLKGSDFVN PSQKESEKVL YTVPLTQGS SGESLDLADA  
 101 NFLEHFQHLF BETTVFGIDQ KLVWSDLDTR NFSQPTQEPD TSNVSEKIS  
 151 SDTKENRKDL ETEDPSKKSQ LKEVSSDLPK SPETAVAAIS EDLEISENIS  
 201 ARDPLQGLAF FYKNTSSQSI SEKDSFQGI IFSGSGANSQ LGFENLKAPK  
 251 SGAAVYSRDR IVFENLVKGL SFISCESLED GSAAGVNIVV THCGDVTLLD  
 301 CATGLDLEAL RLVKDFSRGG AVFTARNHEV QNNLAGGILS VVGKNGAIVV  
 351 EKNSAEKSNG GAFACGSFVY SNNENTALWK ENQALSGGAI SSASDIDIQG  
 401 NCSAIEFSGN QSLIALGEHI GLTDFVGGGA LAAQGTTLTLR NNAVVCVKV  
 451 TSKTHGGAIL AGTVDLNETI SEVAFKQNTA ALTGGALSAN DKVIIANNFG  
 501 EILFEQNEVR NHGGAIFYCG RSNPKLEQKD SGENINIIGN SGAITFLKNK  
 551 ASVLEVMTQA EDYAGGGALW GHNVLLDSNS GNIQFIGNIG GSTFWIGEYV  
 601 GGGAILSTDR VTISNNSGDV VFKGNKGQCL AQKYVAPQET APVESDASST  
 651 NKDEKSLNAC SHGDHYPPKT VEEVPPSLI EEHPVVSSTD IRGGGAILAQ  
 701 HIFITDNTGN LRFSGNLGGG EESSTVGDLA IVGGGALLST NEVNVCNQN  
 751 VVFSNDVTSN GCDSGGAILA KKVDISANHS VEFVSNNGSK FGGAVCALNE  
 801 SVNITDNGSA VSFSKNRTRL GGAGVAAPQG SVTICGNQGN IAFKENFVFG

-81-

5 851 SENQSRGGGA IIANSSVNIQ DNAGDILFVS NSTGSYGGAI FVGSIVASEG  
 901 SNPRTLITIG NSGDILFAKN STQTAASLSE KDSFGGGAIY TQNLKIVKNA  
 951 GNVSFYGNRA PSGAGVQIAD GGTVCLEAFG GDILFEGNIN FDGSFNAILH  
 10 1001 CGNDSKIVEL SAVQDKNIIF QDAITYEENT IRGLPDKDVS PLSAPSLIFN  
 1051 SKPQDDSAQH HEGTIRFSRG VSKIPQIAAI QEGTLALSQN AELWLAGLKQ  
 1101 ETGSSIVLSA GSILRIFDSQ VDSSAPLPTE NKEETLVSAG VQINMSSPTP  
 1151 NKDKAVDTPV LADIISITVD LSSFVPEQDG TLPLPPEIII PKGTKLHSNA  
 1201 IDLKIIDPTN VGYENHALLS SHKDIPLISL KTAEGMTGTP TADASLSNIK  
 1251 IDVSLPSITP ATYGHTGVWS ESKMEDGRLV VGWQPTGYKL NPEKQALVL  
 1301 NNLWSHYTDL RALKQEIFAH HTIAQRMELD FSTNVWGSGL GVVEDCQNIQ  
 1351 EFDGFKHHLT GYALGLDTQL VEDFLIGGCF SQFFGKTESQ SYKAKNDVKS  
 1401 YMGAAYAGIL AGPWLKIGAF VYGNINNDLT TDYGTGLIST GSWIGKGFIA  
 1451 GTSIDYRYIV NPRRFISAIV STVVPFVEAE YVRIDLPEIS EQGKEVRTFQ  
 15 1501 KTRFENVaip FGFALHAYS RGSRAEVNSV QLAYVFDVYR KGPVSLITLK  
 1551 DAAYSWSKSYG VDIPCKAWKA RLSNNTEWNS YLSTYLAFNY EWREDLIAYD  
 1601 FNGGIRIIF\*

A predicted signal peptide is highlighted.

The cp7287 nucleotide sequence <SEQ ID 84> is:

20 1 ATGGTAGCGA AAAAAACAGT ACGATCTTAT AGGTCTTCAT TTTCTCATTC  
 51 CGTAATAGTA GCAATATTGT CAGCAGGCAT TGCTTTTGAA GCACATTCCCT  
 101 TACACAGCTC AGAACTAGAT TTAGGTGTAT TCAATAACA GTTTGAGGAA  
 151 CATTCTGCTC ATGTTGAAGA GGCTCAACA TCTGTTTAA AGGGATCAGA  
 201 TCCTGTAAAT CCCTCTCAGA AAGAATCCGA GAAGGTTTTG TACACTCAAG  
 25 251 GCAAGAGATC CCAAGGAAGC TCTGGAGAGA GTTTGGATCT CGCCGATGCT  
 301 AATTCTTAG AGCATTTTCA GCATCTTTTT GAAGAGACTA CAGTATTTGG  
 351 TATCGATCAA AAGCTGGTTT GGTCAGATTT AGATACTAGG AATTTTTCCC  
 401 AACCCACTCA AGAACCCTGAT ACAAGTAATG CTGTAAGTGA GAAAATCTCC  
 451 TCAGATACCA AAGAGAATAG AAAAGACCTA GAGACTGAAG ATCCTTCAAA  
 501 AAAAAGTGGC CTAAAGAAG TTTCATCAGA TCTCCCTAAA AGTCCTGAAA  
 30 551 CTGCAGTAGC AGCTATTTCT GAAGATCTTG AAATCTCAGA AACATTCTCA  
 601 GCAAGAGATC CTCTTCAGGG TTTAGCATT TTTTATAAAA ATACATCTTC  
 651 TCAGTCTATC TCTGAAAAGG ATTCTTCATT TCAAGGAATT ATCTTTTCTG  
 701 GTTCAGGAGC TAATTCAGGG CTAGGTTTTG AAAATCTTAA GGCGCCGAAA  
 751 TCTGGGGCTG CAGTTTATTC TGATCGAGAT ATGTGTTTTG AAAATCTTGT  
 35 801 TAAAGGATTG AGTTTTATAT CTGTGGAATC TTTAGAAGAT GGCTCTGCCG  
 851 CAGGTGTAAG CATTGTTGTG ACCCATTTGT GTGATGTAAC TCTACTGAT  
 901 TGTGCCACTG GTTAGACCT TGAAGCTTTA CGTCTGGTTA AAGATTTTTC  
 951 TCGTGGAGGA GCTGTTTTCA CTGCTCGCAA CCATGAAGTG CAAAATAACC  
 1001 TTGCAGGTGG AATCTATACC GTTGTAGGCA ATAAAGGAGC TATTGTTGTA  
 40 1051 GAGAAAAATA GTGCTGAGAA GTCCAATGGA GGAGCTTTTG CTGCGGAAG  
 1101 TTTTGTTTAC AGTAACAACG AAAACACCGC CTGTGAGAAA GAAAATCAAG  
 1151 CATTATCAGG AGGAGCCATA TCCTCAGCAA GTGATATTGA TATTCAAGGG  
 1201 AACTGTAGCG CTATTGAATT TTCAGGAAAC CAGTCTCTAA TTGCTCTTGG  
 1251 AGAGCATATA GGGCTTACAG ATTTGTAGG TGGAGGAGCT TTAGCTGCTC  
 45 1301 AAGGGACGCT TACCTTAAGA AATAATGCAG TAGTGAATG TGTAAAAAC  
 1351 ACTTCTAAAA CACATGGTGG AGCTATTTTA GCAGGTACTG TTGATCTCAA  
 1401 CGAAACAATT AGCGAAGTTG CCTTTAAGCA GAATACAGCA GCTCTAACTG  
 1451 GAGGTGCTTT AAGTGCAAAT GATAAGGTTA TAATTGCAA TAACTTTGGA  
 50 1501 GAAATCTTTT TTGAGCAAAA CGAAGTGAGG AATCACGGAG GAGCCATTTA  
 1551 TTGTGGATGT CGATCTAATC CTAAGTTAGA ACAAAGGAT TCTGGAGAGA  
 1601 ACATCAATAT TATTGGAAAC TCCGAGCTA TCACTTTTTT AAAAAATAAG  
 1651 GCTTCTGTTT TAGAAGTGAT GACACAAGCT GAAGATTATG CTGGTGGAGG  
 1701 CGCTTTATGG GGGCATAATG TTCTTCTAGA TTCCAATAGT GGGAAATATTC  
 1751 AATTATAGG AAATATAGGT GGAAGTACCT TCTGGATAGG AGAATATGTC  
 55 1801 GGTGGTGGTG CGATTCTCTC TACTGATAGA GTGACAATTT CTAATAACTC  
 1851 TGGAGATGTT GTTTTTAAAG GAAACAAAGG CCAATGTCTT GCTCAAAAAT  
 1901 ATGTAGCTCC TCAAGAAACA GCTCCCGTGG AATCAGATGC TTCATCTACA  
 1951 AATAAAGACG AGAAGAGCCT TAATGCTTGT AGTCATGGAG ATCATTATCC  
 2001 TCCTAAACT GTAGAAGAGG AAGTGCCACC TTCATTGTGA GAAGAACATC  
 60 2051 CTGTTGTTTC TTCGACAGAT ATTCGTGGTG GTGGGGCCAT TCTAGCTCAA  
 2101 CATATCTTTA TTACAGATAA TACAGGAAAT CTGAGATTCT CTGGGAACCT  
 2151 TGGTGGTGGT GAAGAGTCTT CTACTGTCCG TGATTTAGCT ATCGTAGGAG  
 2201 GAGGTGCTTT GCTTCTACT AATGAAGTTA ATGTTTGCAG TAACCAAAAT  
 2251 GTTGTTTTTT CTGATAACGT GACTTCAAAT GGTGTGATT CAGGGGGAGC  
 65 2301 TATTTTAGCT AAAAAAGTAG ATATCTCCGC GAACCACTCG GTTGAATTTG

2351 TCTCTAATGG TTCAGGGAAA TTCGGTGGTG CCGTTTGC GC TTTAAACGAA  
 2401 TCAGTAAACA TTACGGACAA TGGCTCGGCA GTATCATCTT CTAAAAATAG  
 2451 AACACGTCTT GGCGGTGCTG GAGTTGCAGC TCCTCAAGGC TCTGTAACGA  
 5 2501 TTTGTGGAAA TCAGGGAAAC ATAGCATTTA AAGAGAACTT TGTTTTTGGC  
 2551 TCTGAAAATC AAAGATCAGG TGGAGGAGCT ATCATTGCTA ACTCTTCTGT  
 2601 AAATATTTCAG GATAACGCAG GAGATATCCT ATTTGTAAGT AACTCTACGG  
 2651 GATCTTATGG AGGTGCTATT TTTGTAGGAT CTTTGGTTGC TTCTGAAGGC  
 2701 AGCAACCCAC GAACGCTTAC AATTACAGGC AACAGTGGGG ATATCCTATT  
 10 2751 TGCTAAAAAT AGCACGCAAA CAGCCGCTTC TTTATCAGAA AAAGATTCCCT  
 2801 TTGGTGGAGG GGCCATCTAT ACACAAAACC TCAAAATTGT AAAGAATGCA  
 2851 GGGAACGTTT CTTTCTATGG CAACAGAGCT CCTAGTGGTG CTGGTGTCCA  
 2901 AATTGCAGAC GGAGGAAGTG TTTGTTTAGA GGCTTTTGGG GGAGATATCT  
 2951 TATTTGAAGG GAATATCAAT TTTGATGGGA GTTCAATGC GATTCACCTA  
 3001 TCGCGGAATG ACTCAAAAAT CGTAGAGCTT TCTGCTGTTC AAGATAAAAA  
 15 3051 TATTATTTTC CAAGATGCAA TTACTTATGA AGAGAACACA ATTCGTGGCT  
 3101 TGCCAGATAA AGATGTCAGT CCTTTAAGTG CCCCTTCATT AATTTTAAAC  
 3151 TCCAGCCAC AAGATGACAG CGCTCAACAT CATGAAGGGA CGATACGGTT  
 3201 TTCTCGAGGG GTATCTAAAA TTCCTCAGAT TGCTGCTATA CAAGAGGGAA  
 3251 CCTTAGCTTT ATCACAAAAC GCAGAGCTTT GGTGGCAGG ACTTAAACAG  
 20 3301 GAAACAGGAA GTTCTATCGT ATTGTCTGCG GGATCTATT CCGTATTTT  
 3351 TGATTCCCAG GTTGATAGCA GTGCGCCTCT TCCTACAGAA AATAAAGAGG  
 3401 AGACTCTTGT TTCTGCCGGA GTTCAAATTA ACATGAGCTC TCCTACACCC  
 3451 AATAAAGATA AAGCTGTAGA TACTCCAGTA CTTGCAGATA TCATAAGTAT  
 3501 TACTGTAGAT TTGTCTTCAT TTGTCTCTGA GCAAGACGGA ACTCTTCCTC  
 25 3551 TTCTCTCTGA AATTATCATT CCTAAGGGAA CAAAATTACA TTCTAATGCC  
 3601 ATAGATCTTA AGATTATAGA TCCTACCAAT GTGGGATATG AAAATCATGC  
 3651 TCTTCTAAGT TCTCATAAAG ATATTCCATT AATTTCTCTT AAGACAGCGG  
 3701 AAGGAATGAC AGGGACGCCT ACAGCAGATG CTTCTCTATC TAATATAAAA  
 3751 ATAGATGTAT CTTTACCTTC GATCACACCA GCAACGTATG GTCACACAGG  
 30 3801 AGTTTGGTCT GAAAGTAAA TGGAAAGATGG AAGACTTGTA GTCGGTTGGC  
 3851 AACCTACGGG ATATAAGTTA AATCCTGAGA AGCAAGGGGC TCTAGTTTTG  
 3901 AATAATCTCT GGAGTCATTA TACAGATCTT AGAGCTCTTA AGCAGGAGAT  
 3951 CTTTGCTCAT CATAAGTAG CTCAAAGAAT GGAGTTAGAT TTCTCGACAA  
 4001 ATGTCTGGGG ATCAGGATTA GGTGTGTGTTG AAGATTGTCA GAACATCGGA  
 35 4051 GAGTTTGATG GGTTCAAACA TCATCTCACA GGGTATGCCC TAGGCTTGGG  
 4101 TACACAACCTA GTTGAAGACT TCTTAATTGG AGGATGTTTC TCACAGTTCT  
 4151 TTGGTAAAAC TGAAAGCCAA TCCTACAAAG CTAAGAACGA TGTGAAGAGT  
 4201 TATATGGGAG CTGCTTATGC GGGGATTTTA GCAGGTCCTT GGTAAATAAA  
 4251 AGGAGCTTTT GTTTACGGTA ATATAAACAA CGATTTGACT ACAGATTACG  
 40 4301 GTACTTTAGG TATTTCAACA GGTTCATGGA TAGGAAAAGG GTTTATCGCA  
 4351 GGCACAAGCA TTGATTACCG CTATATTGTA AATCCTCGAC GGTATATATC  
 4401 GGCAATCGTA TCCACAGTGG TTCCTTTTGT AGAAGCCGAG TATGTCCGTA  
 4451 TAGATCTTCC AGAAATTAGC GAACAGGGTA AAGAGGTTAG AACGTTCCAA  
 4501 AAAACTCGTT TTGAGAATGT CGCCATTCCT TTTGGATTG CTTTAGAACA  
 45 4551 TGCTTATTCG CGTGGCTCAC GTGCTGAAGT GAACAGTGTA CAGCTTGCTT  
 4601 ACGTCTTTGA TGTATATCGT AAGGGACCTG TCTCTTTGAT TACACTCAAG  
 4651 GATGCTGCTT ATTCTTGGAA GAGTTATGGG GTAGATATTC CTTGTAAAC  
 4701 TTGGAAGGCT CGCTTGAGCA ATAATACGGA ATGGAATTCA TATTTAAGTA  
 4751 CGTATTTAGC GTTTAATTAT GAATGGAGAG AAGATCTGAT AGCTTATGAC  
 50 4801 TTCAATGGTG GTATCCGTAT TATTTTCTAG

The PSORT algorithm predicts an inner membrane location (0.106).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 42A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 42B) and for FACS analysis (Figure 42C). A his-tagged protein was also expressed.

55 The cp7287 protein was also identified in the 2D-PAGE experiment and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7287 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.



**Example 43**

The following *C.pneumoniae* protein (PID 4377105) was expressed <SEQ ID 85; cp7105>:

```

1  MSLYQKWWNS QLKSLCYST VAALIFMIPS QESFADSLID LNLGLDPSVE
51  CLSGDGAFSV GYFTKAGSTP VEYQPFKYDV SKKTFTILSV ETANQSGYAY
5  101  GISYDGTITV GTCSLGAGKY NGAKWSADGT LTPLTGITGG TSHTEARAIS
151  KDTQVIEGFS YDASGQPKAV QWASGATTVT QLADISGGSR SSYAYAISDD
201  GTIIVGSMES TITRKTAVK WVNNVPTYLG TLGGDASTGL YISGDGTIVV
251  GAANTATVTN GNQESHAYMY KDNQMKD*

```

The cp7105 nucleotide sequence <SEQ ID 86> is:

```

10      1  GTGAGTCTAT ATCAAAATG GTGGAACAGT CAGTTAAAGA AGAGCCTCTG
      51  CTATTGCGACT GTTGCTGCTC TAATATTTAT GATTCCTTCT CAAGAATCCT
101  TTGCAGATAG TCTTATAGAT TTAAATTTAG GTTTAGATCC TTCGGTCGAA
151  TGTCTGTCAG GAGATGGTGC ATTTTCTGTT GGGTATTTTA CTAAGCGGG
201  ATCGACTCCC GTAGAATATC AGCCGTTTAA ATACGACGTA TCTAAGAAGA
15  251  CATTACCAAT CCTTTCGTA GAAACGGCAA ATCAGAGCGG CTATGCTTAC
301  GGAATCTCCT ACGATGGCAC GATCACTGTA GGAACGTGTA GCCTAGGTGC
351  AGGAAAATAT AACGGCGCAA AATGGAGTGC GGATGGCACT TTAACACCCCT
401  TAACTGGAAT CACGGGGGGG ACGTCACATA CGGAACGCGG TCGGATTTCT
451  AAGGATACTC AGGTGATCGA GGGTTTCTCA TATGATGCTT CAGGGCAACC
20  501  CAAGGCTGTG CAGTGGGCAA GCGGAGCGAC TACAGTAACA CAATTAGCAG
551  ATATTTTCAGG AGGCTCTAGA AGCTCTTATG CGTATGCTAT ATCTGATGAT
601  GGCACGATTA TTGTTGGGTC TATGGAGAGC ACGATAACAA GGAAACTAC
651  AGCTGTAAAA TGGGTAAATA ATGTTCTTAC GTATCTGGGA ACCTTAGGAG
701  GAGATGCTTC TACAGTCTT TATATTTCTG GAGACGGCAC CGTGATGTGA
25  751  GGTGCGGCAA ATACAGCAAC TGTAAACCAAT GGAATCAGG AATCCCACGC
801  CTATATGTAT AAAGATAACC AAATGAAAGA TTGA

```

The PSORT algorithm predicts an inner membrane location (0.100).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 43A.

The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 43B) and for FACS analysis (Figure 43C). A his-tagged protein was also expressed.

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7105 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

**35 Example 44**

The following *C.pneumoniae* protein (PID 4376802) was expressed <SEQ ID 87; cp6802>:

```

1  MSNQLQPCIS LGCVSYINSF PLSLQLIKRN DIRCVLAPPA DLLNLLIEGK
51  LDVALTSSLG AISHNLGYVP GFGIAANQRI LSVNLYAAPT FFNSPQPRIA
101  ATLESRSSIG LLKVLCLRLW RIPTPHILRF ITTKVLRQTP ENYDGLLLIG
40  151  DAALQHPVLP GFVTYDLASG WYDLTKLPFV FALLHSTSW KEHPLPNLAM
201  EEALQQFESS PEEVLKEAHQ HTGLPPSLLO EYYALCQYRL GEEHYEFK
251  FREYYGTLYQ QARL*

```

A predicted signal peptide is highlighted.

The cp6802 nucleotide sequence <SEQ ID 88> is:

```

45      1  ATGTCTAACC AACTCCAGCC ATGTATAAGC TTAGGCTGCG TAAGTTATAT
      51  TAATTCCTTT CCGCTGTCCC TACAACATCAT AAAAAGAAAC GATATTCGCT
101  GTGTTCTTGC TCCCCCTGCA GACCTCCTCA ACTTGCTAAT CGAAGGGAAA
151  CTCGATGTTG CTTTGACCTC ATCCCTAGGA GCTATCTCTC ATAACCTGGG
201  GTATGTCCCC GCCTTTGGAA TTGCAGCAAA CCAACGTATC CTCAGTGTA

```

251 ACCTCTATGC AGCTCCCACT TTCTTTAACT CACCGCAACC TCGGATTGCC  
 301 GCAACTTTAG AAAGTCGCTC CTCTATAGGA CTCTTAAAG TGCTTTGTCTG  
 351 TCATCTCTGG CGCATCCCAA CTCCTCATAT CCTAAGATTC ATAACACAA  
 401 AAGTACTCAG ACAAACCCCT GAAAAATTATG ATGGCCTCCT CCTAATCGGA  
 451 GATGCAGCGC TACAACATCC TGTACTTCCT GGATTGTGTA CCTATGACCT  
 501 TGCCTCGGGG TGGTATGATC TTACAAAGCT ACCTTTTGTA TTTGCTCTTC  
 551 TTCTACACAG CACCTCTTGG AAAGAACATC CCCTACCCAA CCTTGCGATG  
 601 GAAGAAGCCC TCCAACAGTT CGAATCTTCA CCCGAAGAAG TCCTTAAAGA  
 651 AGCTCATCAA CATAAGGTC TGCCCCCTTC TCTTCTTCAA GAATACTATG  
 701 CCCTATGCCA GTACCGTCTA GGAGAAGAAC ACTACGAAAG CTTTGAAAAA  
 751 TTCCGGGAAT ATTATGAAC CCTCTACCAA CAAGCCCGAC TGTA

The PSORT algorithm predicts an inner membrane location (0.060).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 44A.

The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 44B) and for FACS analysis (Figure 44C). A his-tagged protein was also expressed.

These experiments show that cp6802 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 45

The following *C.pneumoniae* protein (PID 4376390) was expressed <SEQ ID 89; cp6390>:

1 MVFSYYCMGL FFFSGAIISSC GLLVSLGVGL GLSVLGVLLL LLAGLLLFKI  
 51 QSMLREVPKA PDLLDLEDAS ERLRVKASRS LASLPKEISQ LESYIRSAAN  
 101 DLNTIKTWPH KDQRLVETVS RKLERLAAAQ NYMISELCEI SEILEEEHH  
 151 LILAQESLEW IGKSLFSTFL DMESPLNLSH LSEVRPYLAV NDPRLLEITE  
 201 ESWEVVSFHI NVTSAFKKAQ ILFKNNEHSR MKKKLESVQE LLETFIYKSL  
 251 KRSYRELGCL SEKMRIHDN PLFPWVQDQQ KYAHAKNEFG EIARCLEEFE  
 301 KTFFWLDEEC AISYMDCWDF LNESIQNKKS RVD RDYISTK KIALKDART  
 351 YAKVLLLEENP TTEGKIDLQD AQRAFERQSQ EFYTLHTET KVRLEALQQC  
 401 FSDLREATNV RQVRFTNSEN ANDLKESEFEK IDKERVRYQK EORLYWETID  
 451 RNEQELREEI GESLRQLNRR KGYRAGYDAG RLKGLLRQWK KNLRDVEAHL  
 501 EDATMDFEHE VSKSELCSVR ARLEVLEEL MDMSPKVADI EELLSYBERC  
 551 ILPIRENLER AYLQYNKCSE ILSKAKFFFP EDEQLLVSEA NLREVGAQLK  
 601 QVQKQCQERA QKFAIFEKHI QEOKSLIKEQ VRSFDLAGVG FLKSELLSIA  
 651 CNLYIKAVVK ESIPVDVPCM QLYYSYEDN EAVVRNRLN MTERYQNFKR  
 701 SLNSIQFNGD VLLRDPVYQP EGHETRLKER ELQETTLSCK KLKVAQDRLS  
 751 ELESRLSRR

A predicted signal peptide is highlighted.

The cp6390 nucleotide sequence <SEQ ID 90> is:

1 TTGGTATTCT CATACTATTG CATGGGATTA TTTTTTTTCT CTGGAGCTAT  
 51 TTCTAGTTGT GGTCTTTTAG TGCTCTTAGG AGTTGGTTTA GGACTTAGTG  
 40 101 TTTTAGGAGT ACTTTTACTT CTCTTAGCAG GTCTTTTGCT TTTTAAGATC  
 151 CAAAGTATGC TTCGAGAGGT GCCTAAGGCT CCTGATCTAT TAGATTTAGA  
 201 AGATGCAAGT GAACGGCTTA GAGTAAAGGC TAGCCGTCTT TTAGCAAGCC  
 251 TCCCGAAGGA AATCAGTCAG CTAGAGAGCT ACATTCTGTC TGCAGCTAAT  
 301 GATCTAAATA CAATTAAGAC TTGGCCGCAT AAAGATCAAA GACTCGTCGA  
 45 351 GACCGTGTCA CGAAAATTAG AGCGTCTGGC AGCTGCTCAA AACTATATGA  
 401 TTTCTGAAC TCGCAGATT AGTGAGATT TTAGGAAGA GGAGCATCAT  
 451 CTAATTTTGG CTCAGGAATC TCTAGAATGG ATAGGTAAGA GTCTATTTTC  
 501 TACCTTTCTG GACATGGAAT CTTTTTTTAAA TTTGAGCCAT CTATCTGAAG  
 551 TGCGTCCGTA CTTAGCTGTA AATGATCCTA GATTATTAGA AATTACCGAA  
 50 601 GAATCTTGGG AAGTAGTGAG TCATTTTCATA AATGTAACGT CTGCTTTTAA  
 651 GAAAGCTCAG ATTCTTTTAA AGAACAACGA ACATTCTCGG ATGAAGAAGA  
 701 AAGTAGAAAG TGTTCAGAG TTAAGTGAAG CATTATTTTA TAAGAGTTTA  
 751 AAGAGAAGT ATCGAGAATT AGGATGCTTA AGTGAAAAGA TGAGATCAT  
 801 TCACGACAAT CCTCTCTTCC CTTGGGTGCA AGATCAGCAG AAGTATGCTC  
 55 851 ATGCTAAGAA TGAATTTGGA GAGATTGCGC GGTGTTTAGA GGAGTTTGAA  
 901 AAGACGTTCT TCTGGTTGGA TGAGGAGTGT GCTATTTCTT ACATGGAAGT

```

5      951 TTGGGATTTT CTAAATGAGT CTATTCAGAA TAAGAAGTCC AGAGTAGATC
      1001 GAGATTATAT ATCCACGAAG AAAATTGCAT TAAAGGATAG AGCCCGCACT
      1051 TATGCTAAGG TTCTTTTAGA AGAGAATCCG ACTACAGAGG GTAAAATAGA
      1101 TTTGCAAGAC GCTCAAAGAG CCTTTGAGCG TCAAAGTCAG GAGTTTATA
      1151 CACTAGAGCA TACGGAAACA AAGGTGAGAC TAGAAGCACT TCAACAGTGC
      1201 TTCTCGGATC TTAGGGAGGC GACGAACGTA AGGCAAGTTA GGTTTACAAA
      1251 TTCTGAAAAAT GCGAATGATT TAAAGGAGAG TTTCGAGAAG ATAGATAAAG
      1301 AGCGTGTGCG ATATCAAAAA GAGCAAAGGC TCTATTGGGA AACAATAGAT
      1351 CGCAATGAGC AAGAGCTTAG GGAAGAGATT GGGGAGTCGC TTCGTTTACA
10     1401 AAATCGGAGA AAAGGTATA GGGCTGGATA TGATGCTGGG CGTTTAAAAG
      1451 GTTTGTTGCG TCAGTGAAG AAAAATCTCC GCGATGTGGA AGCCACCTT
      1501 GAAGATGCAA CTATGGATT TTAGCATGAA GTAAGCAAGA GCGAATTGTG
      1551 CAGTGTTCGG GCGAGGCTCG AGGTTCTAGA AGAAGAGCTG ATGGATATGT
      1601 CTCTGAAAGT TGCGGATATA GAAGAGTTGT TGTCTTATGA AGAGCGTTGT
15     1651 ATTCTTCCTA TTAGGGAAAA TTTAGAAAGG GCATACCTCC AATATAATAA
      1701 GTGTTCTGAA ATTTTATCCA AGGCAAAGTT CTCTTTCCG GAAGACGAGC
      1751 AATTGCTAGT TTCGGAAGCG AATCTAAGAG AGGTGGGTGC CCAGTTAAAA
      1801 CAAGTACAGG GAAAAATGTC AGAGAGGGCC CAAAAGTTCT CAATATTGTA
      1851 AAAGCATATT CAGGAGCAGA AAAGCCTTAT TAAAGAGCAA GTGCGGAGTT
20     1901 TTGATCTAGC GGGAGTTGGG TTTTAAAGA GTGAGCTTCT TAGTATTGCT
      1951 TGTAACTTTT ATATAAAGGC GGTGTGTAAG GAGTCTATAC CAGTTGATGT
      2001 GCCTTGTATG CAGTTATATT ATAGTTATTA CGAAGATAAT GAAGCTGTAG
      2051 TCGGAAACCG CCTTTTAAAT ATGACGAGAG GGTATCAAAA TTTTAAAGG
      2101 AGTTTGAATT CCATACAATT TAATGGTGAC GTTCTTTTAC GGGATCCGGT
25     2151 CTATCAACCT GAAGGTCATG AGACCAGGCT AAAGGAACGG GAGCTACAAG
      2201 AAACAACCTT GTCTTGTAAG AAATTAAAAG TGGCTCAAGA TCGTCTTTCT
      2251 GAATTAGAGT CAAGGCTGTC TAGGAGATAG

```

The PSORT algorithm predicts a periplasmic location (0.932).

The protein was expressed in *E. coli* and purified as a GST-fusion product, as shown in Figure 45A.

30 The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 45B) and for FACS analysis (Figure 45C). A his-tagged protein was also expressed.

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

35 These experiments show that cp6390 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 46

The following *C. pneumoniae* protein (PID 4376272) was expressed <SEQ ID 91; cp6272>:

```

40     1  MKRCFLFLAS FVLMGSSADA LTHQEAVKKK NSYLSHFKSV SGIVTIEDGV
      51 LNIHNNLRIQ ANKVYVENTV GQSLKLVAHG NVMVNYRAKT LVCDYLEYYE
      101 DTDSCLLTNG RFAMYPWFLG GSMITLTPEI IVIRKGYIST SEGPKKDLCL
      151 SGDYLEYSSD SLLSIGKTTL RVCRIPIFL PPFSIMPMEI PKPPINFRGG
      201 TGGFLGSYLG MSYSPISRKH FSSTFFLDSF FKHGVGMGFN LHCSQKQVPE
      251 NVFNMKSYA HRLAIDMAEA HDRYRLHGDF CFTHKHVNF SGEYHLSDSWE
      301 TVADIFPNMF MLKNTGPTRV DCTWNDNYFE GYLTSVSVKN SFQNAQQLP
45     351 YLTLRQYPIS IYNTGVYLEN IVECGYLNFA FSDHIVGENF SSLRLAARPK
      401 LHKTVPLPIG TLSSTLGSSL IYSDVPEIS SRHSQLSAKL QLDYRFLHKK
      451 SYIQRRHIE PFVTFITETR PLAKNEDHYI FSIQDAFHSL NLLKAGIDTS
      501 VLSKTNPRFP RIHAKLWTH ILSNTESKPT FPKTACELSL PFGKKNTVSL
      551 DAEWIWKKHC WDHMNIWREW IGNDNVAMTL ESLHRSKYSL IKCDRENFIL
50     601 DVSRPIDQLL DSPLSDHRNL ILGKLFVRPH PCWNYRLSLR YGWHRQDTPN
      651 YLEYQMILGT KIFEHWQLYG VYERREADSR FFFFLKLDKP KKPPF*

```

A predicted signal peptide is highlighted.

The cp6272 nucleotide sequence <SEQ ID 92> is:

```

1  ATGAAACGTT GCTTCTTATT TCTAGCTTCC TTTGTTCTTA TGGGTTCCCTC

```

51 AGCTGATGCT TTGACTCATC AAGAGGCTGT GAAAAAGAAA AACTCCTATC  
 101 TTAGTCACTT TAAGAGTGT TCTGGGATTG TGACCATCGA AGATGGGGTA  
 151 TTGAATATCC ATAACAACCT GCGGATACAA GCCAATAAAG TGTATGTAGA  
 201 AAATACTGTG GGTCAAAGCC TGAAGCTTGT CGCACATGGC AATGTTATGG  
 5 251 TGAAC TATAG GGCAAAAACC CTAGTTTGTG ATTACCTAGA GTATTACGAA  
 301 GATACAGACT CTTGTCTTCT TACTAATGGA AGATTGCGCA TGTATCCTTG  
 351 GTTTCTAGGG GGGTCTATGA TCACTCTAAC CCCAGAAACC ATAGTCATTC  
 401 GGAAGGGATA TATCTCTACC TCCGAGGGTC CCAAAAAGA CCTGTGCCTC  
 451 TCCGGAGATT ACCTGGAATA TTCTTCAGAT AGTCTTCTTT CTATAGGGAA  
 10 501 GACAACATTA AGGGTGTGTC GCATTCCGAT ACTTTTCTTA CCTCCATTTT  
 551 CTATCATGCC TATGGAGATC CCTAAGCCTC CGATAAACTT TCGAGGAGGA  
 601 ACAGGAGGAT TTCTGGGATC CTATTTGGGG ATGAGCTACT CGCCGATTTC  
 651 TAGGAAGCAT TTCTCCTCGA CATTTTCTTT GGATAGCTTT TTCAAGCATG  
 701 GCGTCGGCAT GGGATTCAAC CTCCATTGTT CTCAGAAGCA GGTTCCTGAG  
 15 751 AATGTCTTCA ATATGAAAAG CTATTATGCC CACCGCCTTG CTATCGATAT  
 801 GGCAGAAGCT CATGATCGCT ATCGCCTACA CGGAGATTTC TGCTTCACGC  
 851 ATAAGCATGT AAATTTTCTT GGAGAATACC ATCTCAGCGA TAGTTGGGAA  
 901 ACTGTTGCTG ACATTTTCCC CAACAACCTC ATGTTGAAAA ATACAGGCCC  
 951 CACACGTGTC GATTGCACCT GGAATGACAA CTATTTTGAA GGGTATCTCA  
 20 1001 CCTTCTCTGT TAAGGTAAAC TCTTTCCAAA ATGCCAACCA AGAGCTCCCT  
 1051 TATTTAACAT TAAGGCAGTA CCCGATTCTT ATTTATAATA CGGGAGTGTA  
 1101 CCTTGAAAAC ATCGTAGAAT GTGGGTATTT AAACCTTGCT TTTAGCGATC  
 1151 ATATCGTTGG CGAGAATTTT TCTTCACTAC GTCTTGCTGC GCGCCCTAAG  
 1201 CTCCATAAAA CTGTGCCTCT ACCTATAGGA ACGCTCTCCT CCACCCTAGG  
 25 1251 GAGTTCTCTG ATTTACTATA GCGATGTTCC TGAGATCTCC TCGCGCCATA  
 1301 GTCAGCTTTC CGCGAAGCTA CAACTTGATT ATCGCTTTCT ATTACATAAG  
 1351 TCCTACATTC AAAGACGCCA TATTATAGAG CCGTTCTGTT CCTTCATTAC  
 1401 AGAGACTCGT CCTCTAGCTA AGAATGAAGA TCATTATATC TTTTCTATTTC  
 1451 AAGATGCCTT TCACTCCTTA AACCTTCTGA AAGCGGGTAT AGATACCTCG  
 30 1501 GTAGTGAGTA AGACTAACCC TCGATTCCCG AGAATCCATG CGAAGCTGTG  
 1551 GACTACCCAC ATCTTGAGCA ATACAGAAAG CAAACCCACG TTTCCCAAAA  
 1601 CTGCATGCCA GCTATCTCTA CCTTTTGGAA AGAAAAATAC AGTCTCCTTA  
 1651 GATGCTGAAT GGATTTGGAA AAAGCACTGT TGGGATCACA TGAACATACG  
 1701 TTGGGAGTGG ATCGGAAATG ACAATGTGGC TATGACTCTA GAATCCCTGC  
 35 1751 ATAGAAGCAA ATACAGCCTG ATTAAGTGTG ACAGGGAGAA CTTTATTTTA  
 1801 GATGTCAGCC GTCCCATTTA CCAGCTTTTA GACTCCCTC TCTCTGATCA  
 1851 TAGGAATCTC ATTTTAGGGA AATTATTTGT ACGACCTCAT CCCTGTTGGA  
 1901 ATTACCGCTT ATCCTTACGC TATGGCTGGC ATCGCCAGGA CACTCCGAAC  
 1951 TACCTAGAAT ACCAGATGAT TCTAGGGACG AAGATCTTCG AACATTGGCA  
 40 2001 GCTCTATGGG GTGTATGAAC GCCGAGAAGC AGATAGTCGA TTTTCTTCT  
 2051 TCTTAAAGCT CGACAAACCT AAAAAACCTC CTTCTTAA

The PSORT algorithm predicts an outer membrane location (0.48).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 46A.

The recombinant protein was used to immunise mice, whose sera were used in a Western blot and for  
 45 FACS analysis (Figure 46B). A his-tagged protein was also expressed.

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6272 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## 50 Example 47

The following *C.pneumoniae* protein (PID 4377111) was expressed <SEQ ID 93; cp7111>:

1 MFEAVIADIQ AREILDSRGY PTLHVKVTTT TGSVGEARVP SGASTGKKEA  
 51 LEFRDTSRPR YQKGVLQAV KNVKEILFPL VKGCSVYEQS LIDSLMMDSD  
 101 GSPNKETLGA NAILGVSLAT AHAAAATLRR PLYRYLGGCF ACSLPCPMMN  
 55 151 LINGGMHADN GLEFQEFMIR PIGASSIKEA VNMGADVHT LKKLLHERGL  
 201 STGVGDEGGF APNLASNEEA LELLLLAIK AGFTPGKDIS LALDCAASSF

251 YNVKTGTYDG RHYEEQIAIL SNLCDRYPID SIEDGLAEED YDGWALLTEV  
 301 LGEKVQIVGD DLFVTNP ELI LEGISNGLAN SVLIKPNQIG TLTETVYAIK  
 351 LAQMAGYTTI ISHRSGETTD TTIADLAVAF NAGQIKTGSL SRSERVAKYN  
 401 RLMEIEEELG SEAIPTDSNV FSYEDSEE\*

5 A predicted signal peptide is highlighted.

The cp7111 nucleotide sequence <SEQ ID 94> is:

1 ATGTTTGAAG CTGTCATTGC CGATATCCAG GCTAGGGAAA TCTTGGATTG  
 51 TCGCGGGTAT CCCACTTTAC ATGTTAAAGT AACCACTAGC ACAGGTCTCTG  
 10 101 TTGGAGAAGC TCGGGTTTCT TCAGGAGCAT CCACAGGGAA AAAAGAAGCC  
 151 TTAGAGTTTC GTGATACAGA TTCTCCTCGT TATCAAGGCA AAGGGGTTTT  
 201 GCAAGCTGTA AAAACGTAA AAGAAATTCT TTTTCCCCTC GTCAAGGGAT  
 251 GTAGTGT TTA TGAGCAATCC TTAATTGATT CTCTGATGAT GGATTCTGAC  
 301 GGCTCTCCGA ACAAGAAAC TCTAGGGGCC AATGCTATTT TAGGAGTCTC  
 351 TCTAGCTACA GCACATGCAG CAGCAGCAAC ACTACGCAGA CCTCTGTATC  
 15 401 GTTATTTAGG AGGGTGT TTT CCCTGCAGTC TTCCCTGTCC TATGATGAAT  
 451 CTGATCAATG GAGGCATGCA TGCCGATAAC GGCTTGGAGT TCCAAGAATT  
 501 TATGATCCGT CCTATTGGAG CCTCTTCCAT CAAAGAAGCT GTCAACATGG  
 551 GTGCTGACGT TTTTCATACT TTGAAAAAT TACTCCATGA AAGAGGCTTA  
 601 TCTACTGGAG TGGGTGACGA AGGAGGCTTC GCCCGAATC TTGCTTCTAA  
 20 651 TGAAGAAGCT CTAGAGCTCC TATTGCTGGC TATTGAAAAA GCAGGCTTTA  
 701 CTCCAGGAAA AGATATATCG CTAGCCTTAG ACTGCGCAGC ATCCTCATTC  
 751 TATAACGTAA AACAGGCAC GTATGATGGG AGGCACTATG AAGAGCAAAT  
 801 CGCAATCCTT TCTAATTTAT GTGATCGCTA TCCTATAGAC TCCATAGAAG  
 851 ATGGTCTTGC TGAAGAAGAC TATGACGGGT GGGCCTTGTT AACTGAAGTT  
 25 901 CTTGGAGAAA AAGTACAGAT TGTGGGTGAT GACCTATTTG TTACAAATCC  
 951 GGAATTAATA TTAGAGGGTA TTAGCAATGG ATTAGCGAAC TCTGTGTGA  
 1001 TTAAACCAAA TCAGATAGGG ACGCTTACTG AAACAGTGTA TGCTATCAAG  
 1051 CTTGCGCAAA TGGCTGGCTA TACTACAATT ATTTCTCATC GCTCAGGAGA  
 1101 AACTACGGAC ACTACGATTG CAGATCTTGC TGTTCCTTC AACGCCGTC  
 30 1151 AAATCAAAAC AGGCTCTTTA TCACGTTCTG AGCGTGTGTC AAAATACAAT  
 1201 AGACTCATGG AAATGAAGA AGAGCTTGA TCCGAAGCAA TTTTCACAGA  
 1251 TTCTAATGTA TTTTCTTAC GAGGATTCT GAGGAATAG

The PSORT algorithm predicts an inner membrane location (0.100).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 47A.

35 The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 47B) and for FACS analysis (Figure 47C). A his-tagged protein was also expressed.

The cp7111 protein was also identified in the 2D-PAGE experiment and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

40 These experiments show that cp7111 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 48

The following *C.pneumoniae* protein (PID 4455886) was expressed <SEQ ID 95; cp0010>:

1 MKSQFSWLVL SSTLACFTSC STVFAATAEN IGPSDSFDGS TNTGTYTPKN  
 45 51 TTTGIDYTLT GDITLQNLGD SAALTKGCFS DTTESLSFAG KGYSLSFLNI  
 101 KSSAEGAALS VTDDKNLSLT GFSSLTFLAA PSSVITTPSG KGAVKCGGDL  
 151 TFDNNGTILF KQDYCEENG AISTKNLSLK NSTGSISFEG NKSSATGKKG  
 201 GAICATGTV D ITNNTAPT LF SNNIAEAAGG AINSTGNCTI TGNTSLVFSE  
 251 NSVTATAGNG GALSGDADVT ISGNQSVTF S GNQAVANGGA IYAKKLTLAS  
 301 GGGGVSPFLT IIVQGT TAGN GGAISILAAG EC SL SAEAGD ITFNGNAIVA  
 50 351 TTPQTTRKNS IDIGSTAKIT NLRAISGHSI FFYDPITANT AADSTD TLNL  
 401 NKADAGNSTD YSGSIVFSGE KLSEDEAKVA DNLSTLTKQP VT LTAGNLVL  
 451 KRGVTLDTKG FTQTAGSSVI MDAGTTLKAS TEEVTLTGLS IPVDSLGEK  
 501 KVVIAASAAS KVALSGPIL LLDNQGNAYE NHDLGKTQDF SFVQLSALGT

551 ATTTDVPVAVP TVATPTHYGY QGTWGMTWVD DTASTPKTKT ATLAWTNTGY  
 601 LPNPERQGPL VPNSLWGSFS DIQAIQGVIE RSALTLCSDR GFWAAGVANF  
 651 LDKDKKGEKR KYRHKSGGYA IGGAAQTCSE NLISFAFCQL FGSDKDFLVA  
 701 KNHTDTYAGA FYIQHITECS GFICLLDKL PGWSHKLPLV LEGQLAYSHV  
 751 SNDLKTKYTA YPEVKGSWGN NAFNMMLGAS SHSYPEYLHC FDTYAPYIKL  
 801 NLTYIRQDSF SEKGTEGRSF DDSNLFNLISL PIGVKFEKFS DCNDFSYDLT  
 851 LSYVPDLIRN DPKCTTALVI SGASWETYAN NLARQALQVR AGSHYAFSPM  
 901 FEVLGQFVFE VRGSSRIYNV DLGGKFQF\*

A predicted signal peptide is highlighted.

10 The cp0010 nucleotide sequence <SEQ ID 96> is:

1 ATGAAATCGC AATTTTCCTG GTTAGTGCTC TCTTCGACAT TGGCATGTTT  
 51 TACTAGTTGT TCCACTGTTT TTGCTGCAAC TGCTGAAAAT ATAGGCCCTT  
 101 CTGATAGCTT TGACGGAAGT ACTAACACAG GCACCTATAC TCCTAAAAAT  
 15 151 ACGACTACTG GAATAGACTA TACTCTGACA GGAGATATAA CTCTGCAAAA  
 201 CCTTGGGGAT TCGGCAGCTT TAACGAAGGG TTGTTTTTCT GACACTACGG  
 251 AATCTTTAAG CTTTGCCGGT AAGGGGTACT CACTTTCTTT TTTAAATATT  
 301 AAGTCTAGTG CTGAAGGCGC AGCACTTTCT GTTACAACGT ATAAAAATCT  
 351 GTCGCTAACA GGATTTTCGA GTCTTACTTT CTTAGCGGCC CCATCATCGG  
 401 TAATCACAAC CCCCTCAGGA AAAGGTGCAG TTAATGTGG AGGGGATCTT  
 20 451 ACATTTGATA ACAATGGAAC TATTTTATTT AAACAAGATT ACTGTGAGGA  
 501 AAATGGCGGA GCCATTTCTA CCAAGAATCT TTCTTTGAAA AACAGCACGG  
 551 GATCGATTTC TTTTGAAGGG AATAAATCGA GCGCAACAGG GAAAAAGGT  
 601 GGGGCTATTT GTGCTACTGG TACTGTAGAT ATTACAAATA ATACCGCTCC  
 651 TACCCTCTTC TCGAACAATA TTGCTGAAGC TGCAGGTGGA GCTATAAATA  
 25 701 GCACAGGAAA CTGTACAATT ACAGGGAATA CGTCTCTGT ATTTTCTGAA  
 751 AATAGTGTGA CAGCGACCGC AGGAAATGGA GGAGCTCTTT CTGGAGATGC  
 801 CGATGTTACC ATATCTGGGA ATCAGAGTGT AACTTTCTCA GGAAACCAAG  
 851 CTGTAGCTAA TGGCGGAGCC ATTTATGCTA AGAAGCTTAC ACTGGCTTCC  
 901 GGGGGGGGGG GGGTATCTCC TTTTCTAACA ATAaTAGTCC AAGGTACCAC  
 30 951 TGCAGGTAAT GGTGGAGCCA TTTCTATACT GGCAGCTGGA GAGTGTAGTC  
 1001 TTTCAGCAGA AGCAGGGGAC ATTACCTTCA ATGGGAATGC CATTTGTGCA  
 1051 ACTACACCAC AAATACAAA AAGAAATCTT ATTGACATAG GATCTACTGC  
 1101 AAAGATCACG AATTTACGTG CAATATCTGG GCATAGCATC TTTTCTTACG  
 1151 ATCCGATTAC TGCTAATACG GCTGCGGATT CTACAGATAC TTTAAATCTC  
 35 1201 AATAAGGCTG ATGCAGGTAA TAGTACAGAT TATAGTGGGT CGATTGTTTT  
 1251 TTCTGGTGAA AAGCTCTCTG AAGATGAAGC AAAAGTTGCA GACAACCTCA  
 1301 CTTCTACGCT GAAGCAGCCT GTAACCTTAA CTGCAGGAAA TTTAGTACTT  
 1351 AAACGTGGTG TCACTCTCGA TACGAAAGGC TTTACTCAGA CCGCGGGTTC  
 1401 CTCTGTTATT ATGGATGCGG GCACAACGTT AAAAGCAAGT ACAGAGGAGG  
 40 1451 TCACTTTAAC AGGTCTTTCC ATTCTGTAG ACTCTTTAGG CGAGGGTAAG  
 1501 AAAGTTGTAA TTGCTGCTTC TGCAGCAAGT AAAAATGTAG CCCTTAGTGG  
 1551 TCCGATTCTT CTTTGGGATA ACCAAGGGAA TGCTTATGAA AATCAGCACT  
 1601 TAGGAAAAAC TCAAGACTTT TCATTTGTGC AGCTCTCTGC TCTGGGTACT  
 45 1651 GCAACAAC TA CAGATGTTCC AGCGGTTCCCT ACAGTAGCAA CTCCTACGCA  
 1701 CTATGGGTAT CAAGGTACTT GGGGAATGAC TTGGGTGAT GATACCGCAA  
 1751 GCACTCCAAA GACTAAGACA GCGACATTAG CTTGGACCAA TACAGGCTAC  
 1801 CTTCCGAATC CTGAGCGTCA AGGACCTTTA GTTCCTAATA GCCTTTGGGG  
 1851 ATCTTTTCA GACATCCAAG CGATTCAAGG TGTATAGAG AGAAGTGCTT  
 1901 TGACTCTTTG TTCAGATCGA GGCTTCTGGG CTGCGGGAGT CGCCAATTTT  
 50 1951 TTAGATAAAG ATAAGAAAGG GGAAAAACGC AAATACCGTC ATAAATCTGG  
 2001 TGGATATGCT ATCGGAGGTG CAGCGCAAAC TTGTTCTGAA AACTTAATTA  
 2051 GCTTTGCCCTT TTGCCAATC TTTGGTAGCG ATAAAGATTT CTTAGTCGCT  
 2101 AAAAATCATA CTGATACCTA TGCAGGAGCC TTCTATATCC AACACATTAC  
 2151 AGAATGTAGT GGGTTCATAG GTTGTCTCTT AGATAAACTT CCTGGCTCTT  
 55 2201 GGAGTCATAA ACCCTCTGTT TTAGAAGGGC AGCTCGCTTA TAGCCACGTC  
 2251 AGTAATGATC TGAAGACAAA GTATACTGCG TATCCTGAGG TGAAAGGTTC  
 2301 TTGGGGGAAT AATGCTTTTA ACATGATGTT GGGAGCTTCT TCTCATTCTT  
 2351 ATCCTGAATA CCTGCATTGT TTTGATACCT ATGCTCCATA CATCAAAC TG  
 2401 AATCTGACCT ATATACGTCA GGACAGCTTC TCGGAGAAAG GTACAGAAGG  
 60 2451 AAGATCTTTT GATGACAGCA ACCTCTTCAA TTTATCTTTG CCTATAGGGG  
 2501 TGAAGTTTGA GAAGTTCTCT GATTGTAATG ACTTTTCTTA TGATCTGACT  
 2551 TATCCTATG TTCTGATCT TATCCGCAAT GATCCCAAAT GCACACAGC  
 2601 ACTTGTAATC AGCGGAGCCT CTTGGGAAAC TTATGCCAAT AACTTAGCAC  
 2651 GACAGGCCTT GCAAGTGCGT GCAGGCAGTC ACTACGCCTT CTCTCCTATG  
 65 2701 TTTGAAGTGC TCGGCCAGTT TGTCTTTGAA GTTCGTGGAT CCTCACGGAT

2751 TTATAATGTA GATCTTGGGG GTAAGTCCA ATTCTAG

The PSORT algorithm predicts an outer membrane location (0.922).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 48A.

The recombinant protein was used to immunise mice, whose sera were used in a Western blot

5 (Figure 48B) and for FACS analysis (Figure 48C). A his-tagged protein was also expressed.

The cp0010 protein was also identified in the 2D-PAGE experiment and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp0010 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## 10 Example 49

The following *C.pneumoniae* protein (PID 4376296) was expressed <SEQ ID 97; cp6296>:

1 MEEVSEYLQQ VENQLESCSK RLTKMETFAL GVRLEAKEEI ESIILSDVNV  
51 RFEVLCDIE DMLSRVEIE RMLRMAELPL LPIKEALTKA FVQHNSCKEK  
101 LTKVEPYFKE SPAYLTSEER LQSLNQTQOR AYKESQKVSG LESEVRACRE  
15 151 QLKDQVRQFE TQGVSLIKEE ILFVTSTFRT KFSYHSFRLH VPCMRLYEEY  
201 YDDIDLERTR ARWMAMSERV RDAFQAFQEM LKEGLVEEAQ ALRETEYWLY  
251 REERKSKKKH\*

The cp6296 nucleotide sequence <SEQ ID 98> is:

1 ATGGAGGAGG TGTCTGAGTA TCTTCAGCAA GTAGAAAATC AGTTGGAATC  
20 51 CTGTTCCAAG CGATTAAACCA AGATGGAAC TTTTGCCCTTA GGTGTGAGGT  
101 TGGAAGCTAA AGAAGAGATA GAGTCTATCA TACTTTCCTGA TGTAGTGAAC  
151 CGTTTTGAGG TTTTATGTAG AGATATTGAA GATATGCTAT CTCGAGTCGA  
201 GGAGATAGAG CGGATGTTAC GTATGGCGGA GCTTCCTCTA CTTCTATAA  
251 AAGAAGCGCT TACCAAGGCT TTTGTACAAC ATAACAGCTG TAAAGAGAAG  
25 301 TTAACCAAGG TAGAGCCTTA CTTTAAAGAG AGCCCTGCAT ATCTAACTAG  
351 TGAAGAGCGA TTGCAGAGTT TGAATCAGAC TTTACAACGT GCGTACAAAG  
401 AGTCCCAAAA GGTTCAGGT TTAGAATCGG AAGTGAGAGC CTGTCGAGAG  
451 CAGCTTAAAG ATCAAGTAAG ACAGTTTGAA ACTCAAGGAG TGAGCTTGAT  
501 AAAAGAAGAG ATTCTCTTTG TGAATAGTAC CTTTAGAACT AAATTTAGCT  
30 551 ATCATTTCATT TCGATTACAT GTTCCTTGCA TGAGGTTGTA TGAGGAGTAT  
601 TATGATGACA TTGATCTAGA GAGAACTCGA GCTCGATGGA TGGCGATGTC  
651 TGAGAGGTAT AGAGATGCTT TTCAGGCATT CCAGGAGATG TTGAAGGAAG  
701 GCCTAGTTGA AGAAGCTCAG GCTCTTAGAG AAACCGAGTA CTGGTTATAT  
751 CGAGAGGAGA GAAAGAGTAA AAAGAAACAT TGA

35 The PSORT algorithm predicts a cytoplasmic location (0.523).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 49A.

The recombinant protein was used to immunise mice, whose sera were used in a Western blot

(Figure 49B) and for FACS analysis (Figure 49C). A his-tagged protein was also expressed.

These experiments show that cp6296 is a surface-exposed and immunoaccessible protein, and that it

40 is a useful immunogen. These properties are not evident from the sequence alone.

## Example 50

The following *C.pneumoniae* protein (PID 4376664) was expressed <SEQ ID 99; cp6664>:

1 MVLFHAQASG RNRVKADAIV LPFWHFKDAK NAASFEEFE PSYLPALENF  
51 QGKTGEIELL YSSPKAKEKR IVLLGLGKNE ELTSDVVFQT YATLTRVLRK  
45 101 AKCSTVNIIL PTISELRISA EEFLVGLSSG ILSLNYDYPR YNKVDRNLET

151 PLSKVTVIGI VPKMADAIFR KEAAIFEGVY LTRDLVNRNA DEITPKKLAE  
 201 VALNKGKEFP SIDTKVLGKD AIAKEKMGLL LAVSKGSCVD PHFIVVRYQG  
 251 RPKSKDHTVL IGKGVTFDSG GLDLKPGKSM LTMKEDMAGG ATVLGILSAL  
 301 AVLELPINVT GIIPATENAI DGASYKMGDV YVGMSGLSVE ICSTDAEGRL  
 351 ILADAITYAL KYCKPTRIID FATLTGAMVV SLGEEVAGFF SNNDVLAEDL  
 401 LEASAESESEP LWRLPLVKKY DKTLLHSDIAD MKNLGSNRAG AITAALFLQR  
 451 FLEESSVAWA HLDIAGTAYH EKEEDRYPKY ASGFGVRSIL YYLENSLSK\*

The cp6664 nucleotide sequence <SEQ ID 100> is:

1 GTGGTTTTAT TTCATGCTCA AGCCTCTGGG CGTAATCGTG TTAAGGCAGA  
 51 TGCTATAGTC CTGCCCTTTT GGCATTTTAA GGATGCAAAA AATGCAGCTT  
 101 CTTTTGAAGC CGAGTTTGAA CCCTCGTATC TCCCCGCTTT AGAAAACCTT  
 151 CAAGGAAAAA CCGGGGAGAT TGAACCTCTT TATAGTAGTC CTAAAGCTAA  
 201 GGAAAAACGC ATTGTCCTCT TAGGCTTAGG GAAAAATGAA GAGCTCACCT  
 251 CTGATGTTGT TTTCCAAACC TATGCGACAC TAACTCGTGT CTTACGTAAA  
 15 301 GCAAAGTGTT CCACAGTCAA TATCATCTTA CCTACAATTT CTGAATTGCG  
 351 GCTTTCTGCC GAAGAATTCT TAGTGGGGTT GTCCTCAGGA ATTTTGTGTCAT  
 401 TAAACTATGA CTACCCACGT TATAATAAGG TAGATCGTAA TCTTGAAACT  
 451 CCTCTTTCTA AAGTCACGGT TATCGGTATC GTTCCCAAAA TGGCGGATGC  
 501 TATCTTTTAGG AAAGAAGCAG CCATTTTCGA AGGCGTATAT CTCACTCGAG  
 20 551 ATCTTGTGAA CAGGAATGCT GATGAAATTA CCCCTAAGAA ATTGGCAGAG  
 601 GTTGCTCTGA ATCTGGGAAA AGAGTTCCCT AGTATTGATA CTAAGGTCTT  
 651 GGGAAAAGAT GCCATCGCCA AAGAGAAAAT GGGACTCCTA TTGGCTGTTT  
 701 CCAAGGGTTC TTGTGTGGAT CCACACTTTA TCGTTGTCCG TTATCAAGGA  
 751 CGTCCTAAGT CTAAAGATCA CACCGTCTTG ATAGGGAAAG GGGTCACCTT  
 25 801 TGA CTCTGGA GGTTTAGACC TCAAGCCTGG AAAATCCATG CTTACTATGA  
 851 AAGAAGACAT GGCAGGTGGG GCTACAGTCC TCGGGATTCT CTCGGCGTTA  
 901 GCAGTTTGTAG AGCTTCCTAT AAATGTCACG GGGATCATTC CTGCTACAGA  
 951 GAATGCTATC GATGGCGCCT CCTATAAAAT GGGAGATGTC TATGTAGGAA  
 1001 TGTCGGGGCT TTCTGTTGAG ATTTGTAGTA CCGATGCTGA GGGACGTCTT  
 30 1051 ATCCTCGCTG ATGCGATTAC ATATGCTTTA AAATATTGTA AACCACACG  
 1101 TATTATAGAT TTTGCAACTC TAACAGGAGC TATGGTAGTC TCTCTAGGAG  
 1151 AAGAGGTGTC AGGTTTCTTT TCCAATAACG ATGTTTTAGC TGAAGATCTT  
 1201 TTAGAGGCGT CAGCCGAAAC CTCCGAGCCG TTATGGAGAC TTCCTCTAGT  
 1251 TAAGAAGTAT GATAAAACAT TGCATTCTGA TATTGCTGAT ATGAAAAATC  
 35 1301 TAGGCAGTAA CCGTGCAGGG GCTATTACAG CAGCATTATT CTTGCAGAGA  
 1351 TTTTTGGAAG AATCTTCGGT AGCTTGGGCA CATCTTGATA TTGCAGGTAC  
 1401 TGCATATCAT GAAAAAGAAG AAGACCGTTA TCCAAAATAT GCTTCAGGTT  
 1451 TTGGTGTTTC TTCTATTCTT TATTACTTAG AAAATAGTCT TTCTAAGTAG

The PSORT algorithm predicts an inner membrane location (0.268).

40 The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 50A), as a his-tagged protein, and as a GST/His fusion. The proteins were used to immunise mice, whose sera were used in Western blot Western blot (50B) and FACS (50C) analyses.

The cp6664 protein was also identified in the 2D-PAGE experiment (Cpn0385) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

45 These experiments show that cp6664 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 51

The following *C.pneumoniae* protein (PID 4376696) was expressed <SEQ ID 101; cp6696>:

50 1 MTLIFVIIIV WNAFLIKLC VIMGLQSRLO HCIEVSQNSN FDSQVKQFIY  
 51 ACQDKTLRQS VLKIFRYHPL LKIHDIARAV YLLMALEEGE DLGLSFLNVQ  
 101 QYPSGAVELF SCGGFPWKGL PYPAEHAIEFG LLLLQIAEFY EESQAYVSKM  
 151 SHFQQALFDH QGSVFPWLWS QENSRLLEKE TTLSQSFLFQ LGMQIHPEYS  
 201 LEDPALGFWM QTRSSSAFV AASGCQSSLG AYSSGDVGVI AYGPCSGDIS  
 251 DCYYFGCCGI AKEFVCQKSH QTTEISFLTS TGKPHPRNTG FSYLRDSYVH  
 55 301 LPIRCKITIS DKQYRVHAAL AEATSAMTFS IFCKGKNCQV VDGPRLRSCS



-91-

351 LDSYKGPND IMILGENDAI NIVSASPYME IFALQGKEKF WNADFLINIP  
 401 YKEEGVMLIF EKKVTSEKGR FFTKMN\*

A predicted signal peptide is highlighted.

The cp6696 nucleotide sequence <SEQ ID 102> is:

```

5      1 TTGACTCTAA TTTTGTGTTAT TATTATCGTT TGGTGCAATG CTTTCTGAT
      51 CAAATGTGTC GTGATAATGG GGCTGCAATC CAGGTTACAA CATTGTATAG
     101 AAGTGTCCTCA GAATTCGAAC TTTGATTACAC AAGTAAAACA GTTTATCTAT
     151 GCGTGCCAAG ATAAGACATT AAGGCAGTCT GTACTCAAGA TTTTCCGCTA
     201 CCATCCTTTA CTA AAAAATTC ATGATATTGC TCGGGCCGTC TATCTTTTGA
10    251 TGGCCTTAGA AGAAGGCGAG GATTTAGGCT TAAGCTTTTT AAATGTACAG
     301 CAGTACCCTT CAGGTGCTGT AGAACTGTTT TCTTGTGGGG GATTTCTTGT
     351 GAAAGGATTA CCTTATCCTG CAGAACATGC GGAATTTGGC CTACTCCTGT
     401 TACAGATCGC AGAGTTTTAT GAAGAGAGTC AGGCATACGT CTCTAAAATG
     451 AGTCATTTTC AACAGGCACT CTTTGATCAC CAAGGGAGCG TCTTCCCTC
15    501 TCTCTGGAGC CAGGAGAACT CTCGACTCCT AAAAGAAAAG ACAACTCTTA
     551 GCCAATCGTT TCTCTTCCAA TTAGGAATGC AAATTCACCC AGAATACAGT
     601 CTTGAGGATC CTGCACTAGG GTTCTGGATG CAAAGAACGC GTTCTTCATC
     651 CGCTTTTGTA GCCGCTTCAG GATGTCAAAG TAGCTTGGGA GCGTATTCCT
     701 CAGGGGATGT CGGTGTTATC GCTTATGGAC CTTGCTCTGG AGACATTAGT
20    751 GATTGTTATT ATTTTGGATG TTGTGGAATC GCTAAAGAGT TCGTGTGCCA
     801 AAAATCTCAC CAAACTACAG AGATTTCTTT TCTCACCTCT ACAGGAAAGC
     851 CTCATCCAG AAATACGGGA TTTTCCTACC TTCGAGATTC CTATGTACAT
     901 CTGCCGATCC GCTGTAAGAT CACTATTTCC GACAAGCAAT ATCGCGTGCA
     951 CGCTGCGTTG GCTGAGGCCA CCTCTGCCAT GACGTTTTCT ATTTCTGTGA
25   1001 AGGGGAAGAA TTGTCAGGTT GTTGACGGCC CTCGCTTGC GTCCTGTTC
     1051 CTAGATTCTT ATAAAGGTCC CGGAAACGAC ATTATGATTC TTGGGGAAAA
     1101 TGACGCAATC AACATTGTTT CTGCAAGTCC CTATATGGAA ATTTTGTGCTT
     1151 TGCAAGGCAA AGAAAAATTT TGGAAATGCAG ACTTTTGTAT TAATATTCCT
     1201 TACAAAGAAG AGGGCGTCAT GTTAATTTTT GAAAAAAG TGACCTCTGA
30   1251 GAAAGGAAGA TTCTTTACGA AGATGAATTA A
  
```

The PSORT algorithm predicts an inner membrane location (0.463).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 51A.

The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 51B) and for FACS analysis (Figure 51C). A his-tagged protein was also expressed.

35 This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6696 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 52

40 The following *C.pneumoniae* protein (PID 4376790) was expressed <SEQ ID 103; cp6790>:

```

      1 MSEHKKSSKI IGIDLGTNS CVSVMEGGQA KVITSSEGTR TTPSIVAFKG
     51 NEKLVGIPAK RQAVTNPEKT LGSTKRFIGR KYSEVASEIQ TVPYTVTSGS
    101 KGDAVFEVDG KQYTPPEEIGA QILMKMKETA EAYLGETVTE AVITVPAYFN
    151 DSQRASTKDA GRIAGLDVCR IPEPTAAAL AYGIDKVGDK KIAVFDLGGG
45   201 TFDISILEIG DGVFEVLSTN GDTLLGGDDF DEVIKWMIE EFKKQEGIDL
     251 SKDNMALQRL KDAAEKAKIE LSGVSTEIN QPFITMDAQG PKHLALTLTR
     301 AQFEKLAASL IERTKSPCIK ALSDAKLSAK DIDDVLLVGG MSRMFAVQET
     351 VKELFGKEPN KGVNPDEVVA IGAAIQGGVL GGEVKDVL LLDVIPLSLGIE
     401 TLGGVMTTLV ERNTTIPTQK KQIFSTAADN QPAVTIVVLQ GERPMARDNK
50   451 EIGRFDLTDI PPAPRGHPQI EVSFDIDANG IFHVSADVA SGKEQKIRIE
     501 ASSGLQDEI QRMVRDAEIN KEEDKKRREA SDAKNEADSM IFRAEKAIKD
     551 YKEQIPETLV KEIEERIENV RNALKDDAPI EKIKEVTEDL SKHMOKIGES
     601 MQSQSASAAA SSAANAKGGP NINTEDLKKH SFSTKPPSNN GSSEDHIEEA
  
```

651 DVEIIDNDDK\*

The cp6790 nucleotide sequence &lt;SEQ ID 104&gt; is:

```

      1  ATGAGTGAAC  ACAAAAAATC  AAGCAAAATT  ATAGGTATAG  ACTTAGGCAC
5      51  AACAAACTCC  TGCCTATCTG  TTATGGAAGG  AGGACAAGCT  AAAGTAATTA
      101  CATCATCCGA  AGGAACAAGA  ACCACGCCAT  CGATCGTTGC  CTTCAAAGGT
      151  AATGAGAAAT  TAGTGGGGAT  TCCAGCAAAA  CGTCAAGCAG  TGACAAATCC
      201  AGAAAAAACT  CTCGGCTCTA  CAAAACGCTT  TATTGGCCGT  AAGTACTCTG
      251  AAGTAGCTTC  GGAAATCCAA  ACCGTTCCCT  ATACAGTCAC  CTCCGGATCT
      301  AAAGGTGATG  CCGTTTTTCG  AGTTGATGGC  AAACAATACA  CTCCAGAAGA
10     351  AATTGGCGCA  CAAATCTTAA  TGAAATGAA  AGAGACAGCA  GAAGCTTATC
      401  TAGGCGAAAC  TGTCACAGAA  GCAGTGATCA  CCGTCCCCGC  ATACTTCAAT
      451  GATTCTCAAC  GAGCATCCAC  AAAAGATGCT  GGACGCATG  CAGGTCTAGA
      501  TGTAAACGT  ATCATTCAG  AACCTACCGC  AGCAGCTCTT  GCCTACGGAA
      551  TCGATAAAGT  CGGTGATAAA  AAAATCGCTG  TCTTCGACCT  TGGTGGAGGA
15     601  ACTTTTGATA  TCTCCATCCT  AGAAATCGGT  GATGGCGTCT  TCGAAGTTCT
      651  ATCTACAAAT  GGAGATACTC  TCCTCGGTGG  AGACGACTTT  GATGAAGTCA
      701  TTATCAAATG  GATGATCGAA  GAATTCAAAA  AACAAGAAGG  CATTGATCTT
      751  AGCAAAGATA  ATATGGCCTT  ACAAAGACTT  AAAGATGCTG  CTGAGAAAGC
      801  AAAAAATAGAA  CTTTCAGGAG  TCTCTTCCAC  AGAAATCAAT  CAGCCATTCA
20     851  TCACAATGGA  TGCACAAGGA  CCTAAACACC  TTGCATTGAC  ACTCACACGT
      901  GCGCAATTCG  AGAAACTCGC  AGCCTCTCTA  ATCGAAAGAA  CAAAATCTCC
      951  ATGCATCAAA  GCACTCAGTG  ACGCAAAACT  TTCCGCTAAG  GATATCGATG
     1001  ATGTTCTCTT  AGTTGGAGGT  ATGTCAAGAA  TGCCCGCAGT  GCAAGAAACT
     1051  GTAAAGAAG  TCTTCGGCAA  AGAGCCTAAT  AAAGGAGTCA  ACCCCGACGA
25     1101  AGTTGTTGCT  ATTGGAGCCG  CAATTCAAGG  TGGTGTCTT  GCGGAGAAG
      1151  TTAAGGATGT  TCTACTTCTA  GACGTTATCC  CCCTATCTCT  GGGTATCGAA
      1201  ACTCTAGGAG  GCGTCATGAC  GACTCTGGTA  GAGAGAAATA  CTACAATCCC
      1251  TACACAGAAA  AAACAAATCT  TCTCCACAGC  TGCTGATAAC  CAGCCTGCGG
      1301  TTACCATCGT  AGTTCTCCAA  GGAGAGCGTC  CCATGGCCAA  AGATAACAAG
30     1351  GAAATCGGAA  GATTGATCT  TACAGATATC  CCTCCGGCTC  CTCGAGGCCA
      1401  TCCTCAAATC  GAAGTCTCCT  TCGATATCGA  TGCAAACGGA  ATTTTCCATG
      1451  TCTCAGCTAA  AGATGTTGCC  AGCGGTAAAG  AACAGAAAAT  TCGTATCGAA
      1501  GCAAGCTCAG  GACTTCAAGA  AGATGAAATC  CAAAGAAATG  TTCGAGATGC
      1551  CGAAATTAAT  AAGGAAGAAG  ATAAAAACG  TCGTGAAGCT  TCAGATGCTA
35     1601  AAAATGAAGC  CGATAGCATG  ATCTTCAGAG  CCGAAAAAGC  TATTAAAGAT
      1651  TATAAGGAGC  AAATTCCTGA  AACTTTAGTT  AAAGAAATCG  AAGAGCGAAT
      1701  CGAAAACGTG  CGCAACGCAC  TCAAAGATGA  CGCTCCTATT  GAAAAAATTA
      1751  AAGAGGTTAC  TGAAGACCTA  AGCAAGCATA  TGCAAAAAAT  TGGAGAGTCT
      1801  ATGCAATCGC  AGTCTGCATC  AGCAGCAGCA  TCATCGGCAG  CCAATGCTAA
40     1851  AGGTGGACCT  AACATCAATA  CAGAAGATTT  GAAAAAACAT  AGTTTCAGTA
      1901  CGAAGCCTCC  TTCAAATAAC  GGTTCCTCAG  AAGACCATAT  CGAAGAAGCT
     1951  GATGTAGAAA  TTATTGATAA  CGACGATAAG  TAA

```

The PSORT algorithm predicts an inner membrane location (0.151).

45 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 52A) and a his-tagged product. The proteins were used to immunise mice, whose sera were used in Western blot (Figure 52B) and FACS (Figure 52C) analyses.

The cp6790 protein was also identified in the 2D-PAGE experiment (Cpn0503).

These experiments show that cp6790 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### 50 Example 53

The following *C.pneumoniae* protein (PID 4376878) was expressed <SEQ ID 105; cp6878>:

```

      1  MNVPDSKNLH  PPAYELLEIK  ARITQSYKEA  SAILTAIPDG  ILLSETGHF
5      51  LICNSQAREI  LGIDENLEIL  NRSFTDVLDP  TCLGFSIQEA  LESLKVPKTL
      101  RLSLCKESKE  KEVELFIRKN  EISGYLFIQI  RDRSDYKQLE  NAIRYKNIA
55     151  ELGKMTATLA  HEIRNPLSGI  VGFASILKKE  ISSPRHQRL  SSIISGTRSL
      201  NNLVSSMLEY  TKSQPLNLKI  INLQDFFSSL  IPLLSVSFPN  CKFVREGAQP

```

251 LFRSIDPDRM NSVVWNLVKN AVETGNSPIT LTLHTSGDIS VTNPGTIPSE  
 301 IMDKLFTPPFF TTKREGNGLG LAEAQKIIRL HGGDIQLKTS DSAVSFFIII  
 351 PELLAALPKE RAAS\*

The cp6878 nucleotide sequence <SEQ ID 106> is:

5 1 ATGAACGTCC CTGATTCCAA GAACCTCCAT CCTCCTGCAT ACGAACTCCT  
 51 AGAGATCAAG GCTCGCATCA CACAATCTTA TAAAGAAGCG AGTGCTATAC  
 101 TGACAGCGAT TCCTGATGGT ATCCTATTAC TTTCTGAAAC AGGACACTTT  
 151 CTTATCTGCA ATTCACAAGC ACGTGAAATT CTAGGAATTG ATGAAAATCT  
 10 201 AGAAATCTTT AATAGATCCT TTACCGATGT TCTCCCGCAT ACGTGTCTTG  
 251 GATTTTCTAT TCAAGAGGCT CTTGAATCTC TAAAAGTCCC TAAAACCTCT  
 301 AGACTCTCTC TCTGTAAAGA ATCTAAAGAA AAAGAAGTGG AACTCTTCAT  
 351 CCGTAAAAAC GAGATCAGTG GATACCTGTT TATCCAAATC CGCGATCGGT  
 401 CCGACTATAA ACAACTAGAA AACGCTATAG AAAGATATAA AAATATCGCA  
 451 GAACCTGGGA AAATGACGGC TACCCTAGCT CACGAAATCC GCAATCCGCT  
 15 501 AAGTGAATC GTTGGATTG CCTCTATCCT AAAGAAAGAG ATTTCCTCTC  
 551 CTCGCCACCA ACGAATGCTC TCCTCAATCA TCTCCGGCAC AAGGTCTCTA  
 601 AATAACCTTG TCTCTTCTAT GTTAGAATAT ACAAATCAC AACCCTTGAA  
 651 CCTAAAGATT ATAAATTTAC AAGACTTCTT CTCTTCTCTT ATCCCTCTGC  
 701 TCTCCGTCTC TTTCCCGAAT TGCAAGTTTG TAAGAGAGGG CGCACAACCT  
 20 751 CTATTCAGAT CTATAGATCC TGATCGGATG AACAGTGTG TTTGGAACCT  
 801 AGTGAAAAAT GCTGTAGAAA CAGGGAACCT TCCGATCACT CTGACCCTGC  
 851 ATACATCGGG AGACATCTCG GTAACGAACC CCGGAACGAT TCCTTCCGAG  
 901 ATCATGGACA AGCTCTTCAC TCCATTCTTC ACAACAAAGA GAGAGGGAAA  
 951 TGGTTTGGGA CTGTCTGAAG CTCAAAAAAT TATAAGACTC CATGGAGGAG  
 25 1001 ATATCCAATT AAAACAAGC GACTCCGCCG TTAGCTTCTT CATAATCATC  
 1051 CCCGAACCTC TAGCGGCCCT ACCCAAAGAA AGAGCCGCTA G

The PSORT algorithm predicts an inner membrane location (0.204).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 53A) and as a GST-fusion product. The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 53B) and for FACS analysis.

These experiments show that cp6878 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 54

The following *C.pneumoniae* protein (PID 4377224) was expressed <SEQ ID 107; cp7224>:

35 1 MMKIRKVAL AVGSGGGHIV PALSVKEAFS REGIDVLLLG KGLKNHPSLQ  
 51 QGISYREIPS GLPTVLNPIK IMSRTLSLCS GYLKARKELK IFDPDLVIGF  
 101 GSYHSLPVLL AGLSHKIPLF LHEQNLVPGK VNQLFSRYAR GIGVNFSPVT  
 151 KHFRCPAEV FLPKRSFSLG SPMKRCCTNH TPTICVVGGS QGAQILNTCV  
 201 PQALVKLVNK YPNMYVHHIV GPKSDVMKVQ HVYNRGEVLC CVKPFEEQLL  
 40 251 DVLLAADLVI SRAGATILEE ILWAKVPGIL IPYPGAYGHQ EVNAKFFVDV  
 301 LEGGTMILEK ELTEKLLVEK VTFALDSHNR EKQRNSLAAY SQQRSTKTFH  
 351 AFICECL\*

The cp7224 nucleotide sequence <SEQ ID 108> is:

45 1 ATGATGAAGA AAATTCGAAA AGTAGCCTTG GCTGTAGGAG GTTCAGGAGG  
 51 CCACATTGTC CCAGCTCTCT CGGTAAAGGA AGCTTTTCTT CGTGAAGGAA  
 101 TAGACGTATT ACTACTAGGG AAAGGTCTCA AGAACCATCC TTCTTTGCAA  
 151 CAGGGAATCA GCTATCGGGA AATCCCCTCA GGACTTCCTA CAGTCCCTAA  
 201 TCCCATAAAG ATCATGAGCA GGACCCCTTC TCTATGTTCA GGATACCTGA  
 251 AAGCAAGAAA GGAACCTAAA ATTTTGTACC CTGACCTGGT CATAGGATTT  
 301 GGGAGCTACC ACTCTCTTCC CGTGTGCTC GCAGGACTGT CCCATAAAAT  
 351 TCCCTTATTT CTACACGAAC AAAATCTAGT TCCTGGAAAA GTAAATCAAT  
 401 TGTTTTCCTG CTATGCTCGA GGTATTGGAG TGAATTTCTC CCCCGTTACT  
 451 AAACACTTCC GCTGCCCGC AGAAGAGGTC TTCCTTCCTA AACGAAGCTT  
 501 CTCTTAGGA AGCCCTATGA TGAAGCGATG TACAAATCAT ACCCCTACAA  
 55 551 TCTGTGTTGT TGGAGGTTCT CAGGGAGCAC AGATATTTAA TACTGTGTGT  
 601 CCCCAAGCTC TTGTCAAGCT AGTCAATAAG TACCCTAATA TGTACGTCCA

```

651 TCATATTGTA GGACCTAAAA GTGATGTTAT GAAGGTGCAA CATGTTTACA
701 ATCGTGGAGA GGTCTCTGTC TGTGTGAAGC CGTTCGAAGA GCAACTCCTA
751 GATGTC TTGC CCGCAGA TTTGGTCATC AGTAGGGCAG GAGCCACAAT
801 TTTAGAAGAA ATTCTTTGGG CAAAAGTTCC CGGAATTTTA ATTCCCTATC
851 CAGGAGCTTA TGGACATCAG GAAGTTAATG CTAAATTCTT TGTAGACGTC
901 TTAGAAGGGG GAACTATGAT CCTAGAAAAA GAATTAACAG AGAAGCTATT
951 AGTAGAAAAA GTAACGTTTG CTTTAGACTC CCATAACAGA GAAAAACAAC
1001 GCAATTCCTT AGCGGCGTAT AGTCAGCAAA GGTCAACAAA AACATTCCAT
1051 GCATTCAATT GTGAATGCTT ATAG

```

10 The PSORT algorithm predicts an inner membrane location (0.164).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 54A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 54B) and for FACS analysis (Figure 54C). A his-tagged protein was also expressed.

15 This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7224 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 55

The following *C.pneumoniae* protein (PID 4377140) was expressed <SEQ ID 109; cp7140>:

```

20      1  MVRRSISFCL FFLMTLLCCT SCNSRSLIVH GLPGREANEI VVLLVSKGVA
      51  AQKLFOAAAA TAGAATEQMW DIAVPSAQIT EALAILNQAG LPRMKGTSLL
101  DLFAKQGLVP SELQEKIRYQ EGLSEQMAST IRKMDGVVDA SVQISFTTEN
151  EDNLPLTASV YIKHRGVLDN PNSIMVSKIK RLIASAVPGL VPENVSVVSD
201  RAAYSDITIN GPWGLTEEID YVSVWGIILA KSSLTKFRLI FYVLILILFV
25  251  ISCGLLWVIW KTHTLIMTMG GTKGFFNPTP YTKNALEAKK AEGAAADKEK
301  KEDADSQGES KNAETSDKDS SDKDAPEGSN EIEGA*

```

A predicted signal peptide is highlighted.

The cp7140 nucleotide sequence <SEQ ID 110> is:

```

30      1  ATGGTTCGTC GATCTATTTT TTTTGCTTG TTCTTTCTAA TGACATTGCT
      51  GTGCTGTACA AGCTGTAACA GCAGGTCTCT AATTGTGCAC GGTCTTCCTG
101  GCAGAGAAGC GAATGAGATT GTGGTGCTTT TGGTAAGCAA AGGGGTGGCT
151  GCACAAAAAT TGCCTCAAGC TGCAGCGGCT ACAGCCGGAG CAGCTACTGA
201  GCAAATGTGG GATATCGCGG TTCCGTCAGC ACAAATCACA GAGGCCCTTG
25  251  CCATTCTAAA TCAAGCGGGT CTTCACGTA TGAAAGGGAC AAGCCTGTTA
35  301  GATCTTTTTT CAAAACAAGG TCTTGTTCC TCCGAGCTTC AGGAAAAAAT
      351  CCGTTATCAA GAAGGCTTAT CAGAACAGAT GGCCTCTACG ATTAGAAAAA
      401  TGGATGGCGT TGTCGATTGC TCAGTACAGA TTTCCTTCAC TACAGAAAAA
      451  GAAGATAATC TTCCTTTAAC AGCCTCTGTG TATATTAAGC ATCGAGGGGT
501  TTTGGACAAT CCGAACAGCA TTATGGTTTC CAAAATTAAG CGCCTTATTG
40  551  CAAGTGCTGT TCCAGGACTT GTGCCAGAGA ACGTCTCTGT AGTGAGCGAT
      601  CGCGCAGCTT ATAGTGATAT TACAATTAAT GGTCTTTGGG GATTAACAGA
      651  AGAAATCGAT TATGTTTCTG TTTGGGGTAT TATTCTTGCG AAGTCTTCGC
      701  TCACCAAATT CCGTCTCATT TTTTATGTCT TGATTCTCAT TTTATTTGTT
      751  ATTTCTTGTG GTCTCCTTTG GGTCAATTTG AAAACTCATA CTCTCATTAT
45  801  GACTATGGGA GGTACAAAAG GGTCTTTCAA CCCTACACCA TATACAAAGA
      851  ATGCCTTGGG AGCCAAGAAA GCCGAGGGAG CAGCTGCTGA CAAAGAGAAA
      901  AAAGAAGATG CAGATTCACA GGGGGAAGC AAAAATGCGG AAACCAGTGA
      951  TAAAGACTCT AGTGATAAAG ATGCTCCAGA AGGAAGCAAT GAAATTGAGG
1001 GTGCTTAG

```

50 The PSORT algorithm predicts an inner membrane location (0.650).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 55A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 55B) and for FACS analysis (Figure 55C). A his-tagged protein was also expressed.

These experiments show that cp7140 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 56

The following *C.pneumoniae* protein (PID 4377306) was expressed <SEQ ID 111; cp7306>:

```

1  MITKQLRSWL AVLVGSSLLA LPLSGQAVGK KESRVSELPQ DVLLKEISGG
51 FSKVATKATP AVVYIESFPK SQAVTHPSPG RRGPYENPFD YFNDEFFNRF
101 FGLPSQREKP QSKEAVRGTG FLVSPDGYIV TNNHVVEDTG KIHVTLHDGQ
151 KYPATVIGLD PKTDLAVIKI KSQNLPLYLSF GNSDHLKVG D WAIAIGNPFG
201 LQATVTVGVI SAKGRNQLHI ADFEDFIQTD AAINPGNSGG PLLNIDGQVI
251 GVNIAIVSGS GGYIGIGFAI PSLMANRIID QLIRDGQVTR GFLGVTLQPI
301 DAELAACYKL EKVYGALVTD VVKGSPADKA GLKQEDVIIA YNGKEVDSL S
351 MFRNAVSLMN PDTRIVLKVV REGKVIEIPV TVSQAPKEDG MSALQRVGIR
401 VQNLTPETAK KLGIAPETKG ILIISVEPGS VAASSGIAPG QLILAVNRQK
451 VSSIEDLNRT LKDSNNENIL LMVSQGDVIR FIALKPPE*
```

A predicted signal peptide is highlighted.

The cp7306 nucleotide sequence <SEQ ID 112> is:

```

1  ATGATAACTA AGCAATTGCG TTCGTGGCTA GCTGTACTTG TTGGTTCAAG
51 TCTGCTAGCT CTTCTTTTAT CAGGGCAAGC TGTCGGGAAA AAAGAATCTC
101 GAGTTTCCGA GCTGCCTCAA GACGTTCTTC TTAAAGAGAT CTCGGGAGGG
151 TTTTCTAAGG TCGCTACCAA GCGGACTCCC GCTGTTGTGT ACATAGAAAAG
201 TTTCCCAAAG AGCCAGGCTG TAACACATCC TTCTCCTGGA CGCCGTGGGC
251 CTTATGAAAA TCCTTTTGAT TATTTTAATG ATGAGTTTTT CAATCGTTTT
301 TTTGGTCTAC CTTACAGAG GAAAAACCT CAAAGTAAAG AGGCGGTTCTG
351 AGGAACAGGT TTCCTAGTAT CTCCAGATGG CTATATTGTG ACTAATAACC
401 ATGTTGTCGA AGATACAGGT AAGATTACAG TAACCTCTCA TGATGGGCAA
451 AAGTACCCAG CAACTGTAAT CGGACTCGAT CCTAAAACAG ACCTTGCAGT
501 CATTAAAAAT AAATCCCAAA ACCTCCCGTA TCTTCTTTT GGAAACTCCG
551 ACCACTTAAA AGTCGGAGAT TGGGCAATTG CAATTGGAAA TCCCTTCGGT
601 CTTCAAGCTA CGGTCACCGT AGGTGTCATC AGTGCTAAAG GAAGAAATCA
651 ACTCCACATT GCAGATTTTG AAGATTTTAT TCAGACAGAT GCTGCGATTA
701 ATCCAGGCAA CTCTGGAGGC CCTCTTCTAA ATATTGATGG ACAGGTCATC
751 GGTGTTAATA CTGCCATTGT CAGTGGTAGT GGTGGCTATA TTGGAATCGG
801 GTTTGCGATT CCTAGCCTTA TGGCAATAG AATCATAGAT CAGCTGATTC
851 GTGATGGTCA AGTTACCCGA GGATTCTTAG GAGTGACTTT ACAACCTATA
901 GATGCGGAAC TCGCTGCTTG CTACAACTC GAAAAGGTTT ATGGCGCTTT
951 AGTCACAGAT GTTGTTAAAG GATCTCCAGC AGATAAAGCA GGGCTAAAAC
1001 AAGAAGATGT GATCATTGCT TATAATGGGA AAGAAGTCGA TTCACTGAGT
1051 ATGTTCCGTA ATGCTGTTTC TTTAATGAAT CCAGATACAC GTATTGTTCT
1101 AAAGGTAGTT CGTGAAGGAA AGGTTATCGA AATACCCGTG ACAGTTTCTC
1151 AAGCTCCAAA AGAAGATGGA ATGTCGGCTT TACAGCGTGT GGAATCCGT
1201 GTGCAAAACC TAACTCCTGA AACTGCTAAG AAGCTGGGAA TTGCTCCAGA
1251 GACTAAAGGC ATTTTGATTA TAAGTGTTGA ACCAGGGTCT GTAGCAGCTT
1301 CTTCAGGAAT TGCTCCTGGT CAGCTGATCC TTGCTGTGAA TAGACAAAAA
1351 GTATCTTCGA TTGAAGATCT GAATAGAACG TTAAAAGATT CTAACAATGA
1401 GAATATTCTT CTTATGGTTT CTCAAGGAGA TGTTATTTCG TTCATTGCCC
1451 TGAAACCTGA AGAATAA
```

The PSORT algorithm predicts a periplasmic location (0.923).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 56A) and as a GST-fusion product (Figure 56B). The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 56C) and for FACS (Figure 56D) analyses.

The cp7306 protein was also identified in the 2D-PAGE experiment (Cpn0979) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7306 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## 5 Example 57

The following *C.pneumoniae* protein (PID 4377132) was expressed <SEQ ID 113; cp7132>:

```

1  MCNSIAMKKQ KRGFVLMELL MSFTLIALLL GTLGFWYRKI YTVQKQKERI
51  YNFYIEESRA YKQLRTLFSM SLSSSYEEPG SLFSLIFDRG VYRDPKLAGA
101 VRASLHHDTK DQRLELRICN IKDQSYFETQ RLLSHVTHV V LSFQRNPDPE
151 KLPETIALTI TREPKAYPPR TLTYQFAVGK*

```

A predicted signal peptide is highlighted.

The cp7132 nucleotide sequence <SEQ ID 114> is:

```

1  ATGTGTAACT CTATAGCTAT GAAAAAGCAA AAGCGTGGCT TTGTGCTTAT
15 51  GGAATTACTC ATGTCGTTCA CTCTAATTGC TTTGTTATTA GGGACTTTAG
101 GATTTTGGTA TCGGAAAATT TATACTGTAC AAAAGCAAAA AGAACGTATT
151 TATAACTTTT ATATCGAAGA AAGCCGAGCC TACAAGCAGC TCAGAACCCT
201 GTTTAGCATG TCCTTGCTCT CATCTTACGA GGAGCCTGGA TCATTATTTT
251 CTTTAATCTT TGATCGGGGT GTTATCGAG ATCCTAAGCT GGCAGGTGCG
301 GTACGAGCTT CTCTCCATCA TGACACCAAG GATCAGAGAT TGGAACCTCG
20 351 TATTTGTAAT ATTAAGGATC AGTCTTACTT TGAAACACAG CGACTGCTCT
401 CCCACGTGAC CCATGTTGTA CTTTCCTTCC AGAGAAATCC TGATCCTGAA
451 AAACCTTCCTG AAACAATTGC TTTAACTATA ACACGGGAAC CTAAAGCATA
501 TCCTCCAAGG ACGTTAACAT ACCAATTTCG GGTGGGAAA TAA

```

The PSORT algorithm predicts a periplasmic location (0.915).

25 The protein was expressed in *E.coli* and purified as a his-tag product (Figure 57A) or as a GST-fusion. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 57B) and FACS (Figure 57C) analyses.

These experiments show that cp7132 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## 30 Example 58

The following *C.pneumoniae* protein (PID 4376733) was expressed <SEQ ID 115; cp6733>:

```

1  MKTSIPWVLV SSVLAFSCHL QSLANEELLS PDDSFNGNID SGTFTPKTSA
51  TTYSLTGDFV FYEPGKGTP L SDSCFKQTTD NLTF LGNGHS LTFGFDAGT
35 101 HAGAAASTTA NKNLTFSGFS LLSFDSSPST TVTTGQGTLS SAGGVNLENI
151 RKLVVAGNFS TADGGAIKGA SP LLTGTS GD ALFSNNSSST KGGAIATTAG
201 ARIANNTGYV RFLSNIAS TS GGAI DDEG TS ILSNNKFLYF EGNAAKTTGG
251 AICNTKASGS PELIISNNKT LIFASNVAET SGGAIHAKKL ALSSGGFTEF
301 LRNNVSSATP KGGAISIDAS GELSLSAETG NITFVRNTLT TTGSTDTPKR
351 NAINIGSNGK FTELRAAKNH TIFFYDPITS EGTSSDVLKI NNGSAGALNP
40 401 YQGTILFSGE TLTADELKVA DNLKSSFTQP VSLSGGKLLL QKGVLTLEST
451 FSQEAGSLLG MDSGTTLS TT AGSITITNLG INVDSLGLKQ PVSLTAKGAS
501 NKVIVSGKLN LIDIEGNIYE SHMPSHDQLF SLLKITVDAD VDTNVDISSL
551 IPVPAEDPNS EYGFQGWNV NWTDTATNT KEATATWTKT GFVPSPERKS
601 ALVCNTLWGV FTDIRSLQQL VEIGATGMEH KQGFVSSMT NFLHKTGDEN
45 651 RKGFRTSGG YVIGGSAHTP KDDLFTFAFC HLFARDKDCF IAHNNSRTYG
701 GTLFFKHSHT LQPQNYLR LG RAKFSESAIE KFPREIPLAL DVQVSFSSHSD
751 NRMETHYTS L PESEGSWSNE CIAGGIGLDL PFVLSNPHPL PKTFIPQMKV
801 EMVYVSQNSF FESSSDGRGF SIGRLLNLSI PVGAKFVQGD IGDSYTYDLS

```

-97-

851 GFFVSDVYRN NPQSTATLVM SPDSWKIRGG NLSRQAFLLR GSNMYVYNSN  
 901 CELFGHYAME LRGSSRNYNV DVGTKLRF\*

A predicted signal peptide is highlighted.

The cp6733 nucleotide sequence <SEQ ID 116> is:

```

5      1  ATGAAGACTT CGATTCCCTG GGTTTTAGTT TCCTCCGTGT TAGCTTTCTC
      51  ATGTCACCTA CAGTCAC TAG CTAACGAGGA ACTTTTATCA CCTGATGATA
     101  GCTTTAATGG AAATATCGAT TCAGGAACGT TTA CTCCAAA AACTTCAGCC
     151  ACAACATATT CTCTAACAGG AGATGTCCTC TTTTACGAGC CTGGAAAAGG
     201  CACTCCCTTA TCTGACAGTT GTTTTAAGCA AACCACGGAC AATCTTACCT
    10  251  TCTTGGGGAA CGGTCATAGC TTAACGTTTG GCTTTATAGA TGCTGGCACT
     301  CATGCAGGTG CTGCTGCATC TACAACAGCA AATAAGAATC TTACCTTCTC
     351  AGGGTTTTC TTA CTGAGTT TGTATTCTC TCCTAGCACA ACGGTTACTA
     401  CAGGTCAGGG AACGCTTTC TCAGCAGGAG GCGTAAATTT AGAAAATATT
     451  CGTAAACTTG TAGTTGCTGG GAATTTTCT ACTGCAGATG GTGGAGCTAT
    15  501  CAAAGGAGCG TCTTTCCTTT TAACTGGCAC TTCTGGAGAT GCTCTTTTTA
     551  GTAACAACCTC TTCATCAACA AAGGGAGGAG CAATTGCTAC TACAGCAGGC
     601  GCTCGCATAG CAAATAACAC AGGTTATGTT AGATTCTCTAT CTAACATAGC
     651  GTCTACGTCA GGAGGCGCTA TCGATGATGA AGGCACGTCG ATACTATCGA
     701  ACAACAAATT TCTATATTTT GAAGGGAATG CAGCGAAAAC TACTGGCGGT
    20  751  GCGATCTGCA ACACCAAGGC GAGTGGATCT CCTGAAC TGA TAATCTCTAA
     801  CAATAAGACT CTGATCTTTG CTTCAAACGT AGCAGAAACA AGCGGTGGCG
     851  CCATCCATGC TAAAAAGCTA GCCCTTTCCT CTGGAGGCTT TACAGAGTTT
     901  CTACGAAATA ATGTCTCATC AGCAACTCCT AAGGGGGGTG CTATCAGCAT
     951  CGATGCCTCA GGAGAGCTCA GTCTTCTGTC AGAGACAGGA AACATTACCT
    25 1001  TTGTAAGAAA TACCCTTACA ACAACCGGAA GTACCGATAC TCCTAAACGT
    1051  AATGCGATCA ACATAGGAAG TAACGGGAAA TTCACGGAAT TACGGGCTGC
    1101  TAAAAATCAT ACAATTTTCT TCTATGATCC CATCACTTCA GAAGGAACCT
    1151  CATCAGACGT ATTGAAGATA AATAACGGCT CTGCGGGAGC TCTCAATCCA
    1201  TATCAAGGAA CGATCTTATT TTCTGGAGAA ACCCTAACAG CAGATGAACT
    30 1251  TAAAGTTGCT GACAATTTAA AATCTTCATT CACGCAGCCA GTCTCCCTAT
    1301  CCGGAGGAAA GTTATTGCTA CAAAAGGGAG TCACTTTAGA GAGCACGAGC
    1351  TTCTCTCAAG AGGCCGGTTC TCTCCTCGGC ATGGATT CAG GAACGACATT
    1401  ATCAACTACA GCTGGGAGTA TTACAATCAC GAACCTAGGA ATCAATGTTG
    1451  ACTCCTTAGG TCTTAAGCAG CCCGTCAGCC TAACAGCAA AGGTGCTTCA
    35 1501  AATAAAGTGA TCGTATCTGG GAAGCTCAAC CTGATTGATA TTGAAGGGAA
    1551  CATTTATGAA AGTCATATGT TCAGCCATGA CCAGCTCTTC TCTCTATTAA
    1601  AAATCACCGT TGATGCTGAT GTTGATACTA ACGTTGACAT CAGCAGCCTT
    1651  ATCCCTGTTC CTGCTGAGGA TCCTAATTCA GAATACGGAT TCCAAGGACA
    1701  ATGGAATGTT AATTGGACTA CGGATACAGC TACAAATACA AAAGAGGCCA
    40 1751  CGGCAACTTG GACCAAAACA GGATTTGTTC CCAGCCCCGA AAGAAAATCT
    1801  GCGTTAGTAT GCAATACCTT ATGGGGAGTC TTTACTGACA TTCGCTCTCT
    1851  GCAACAGCTT GTAGAGATCG GCGCAACTGG TATGGAACAC AAACAAGGTT
    1901  TCTGGGTTTC CTCCATGACG AACTTCTGTC ATAAGACTGG AGATGAAAAT
    1951  CGCAAAGGCT TCCGTCATAC CTCTGGAGGC TACGTCATCG GTGGAAGTGC
    45 2001  TCACACTCCT AAAGACGACC TATTTACCTT TGGTTCCTGC CATCTCTTTG
    2051  CTAGAGACAA AGATTGTTTT ATCGCTCACA ACAACTCTAG AACCTACGGT
    2101  GGAACTTTAT TCTTCAAGCA CTCTCATACC CTACAACCCC AAAACTATTT
    2151  GAGATTAGGA AGAGCAAAGT TTTCTGAATC AGCTATAGAA AAATTCCCTA
    2201  GGGAAATTCC CCTAGCCTTG GATGTCCAAG TTTCGTT CAG CCATT CAGAC
    50 2251  AACCGTATGG AAACGCACTA TACCTCAT TG CCAGAATCCG AAGGTTCTTG
    2301  GAGCAACGAG TGTATAGCTG GTGGTATCGG CCTAGACCTT CCTTTTGTTC
    2351  TTTCCAACCC ACATCCTCTT TTCAAGACCT TCATTCCACA GATGAAAGTC
    2401  GAAATGGTTT ATGTATCACA AAATAGCTTC TTCGAAAGCT CTAGTGATGG
    2451  CCGTGGTTT AGTATTGGAA GGCTGCTTAA CCTCTCGATT CCTGTGGGTG
    55 2501  CGAAATTCGT GCAGGGGAT ATCGGAGATT CCTACACCTA TGATCTCTCA
    2551  GGATCTTTG TTTCGATGT CTATCGTAAC AATCCCCAAT CTACAGCGAC
    2601  TCTTGTGATG AGCCCAGACT CTTGGAAAAT TCGCGGTGGC AATCTTTCAA
    2651  GACAGGCATT TTTACTGAGG GGTAGCAACA ACTACGCTA CAACTCCAAT
    2701  TGTGAGCTCT TCGGACATTA CGCTATGGAA CTCCGTGGAT CTTCAGGAA
    60 2751  CTACAATGTA GATGTTGGTA CCAAACCTCCG ATTCTAG
  
```

The PSORT algorithm predicts an outer membrane location (0.924).

The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 58A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 58B) and for FACS (Figure 58C) analyses. A GST-fusion protein was also expressed.

The cp6733 protein was also identified in the 2D-PAGE experiment (Cpn0451).

- 5 These experiments show that cp6733 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 59

The following *C.pneumoniae* protein (PID 4376814) was expressed <SEQ ID 117; cp6814>:

```

10      1  MHDALLSILA IQELDIKMIR LMRVKKEHQK ELAKVQSLKS DIRRKVQEKE
      51  LEMENLKTQI RDGENRIQEI SEQINKLENQ QAAVKKMDDEF NALTQEMTTA
     101  NKERRSLEHQ LSDLMDKQAG GEDLIVSLKE SLASTENSSS VIEKEIFESI
     151  KKINEEGKAL LEQRTTELKHA TNPELLSIYE RLLNNKKDRV VVPIENRVCS
     201  GCHIVLTPQH ENLVRKKDRL IFCEHCSRIL YWQESQVNAQ ENSTAKRRRR
     251  RAAV*
```

- 15 The cp6814 nucleotide sequence <SEQ ID 118> is:

```

      1  ATGCATGACG CACTTCTAAG CATTTTGGCT ATTCAAGAGC TTGATATTAA
      51  AATGATTTCG CTTATGCGCG TAAAGAAAGA ACATCAGAAA GAATTGGCTA
     101  AAGTCCAATC TTTAAAAAGT GATATTCTGA GAAAAGTTCA GGAAAAAGAA
     151  CTCGAAATGG AGAATTTGAA AACTCAAATT CGAGATGGAG AGAATCGCAT
     20  201  CCAAGAGATT TCTGAACAAA TCAATAAATT AGAAAATCAG CAAGCTGCTG
     251  TAAAAAAAAT GGATGAGTTT AACGCTCTTA CCCAAGAAAT GACTACAGCA
     301  AACAAAGAAC GTCGCTCTTT AGAGACCCAG CTTAGCGATC TCATGGATAA
     351  GCAAGCTGGA GGCGAAGACC TTATTGTCTC TCTAAAAGAA AGCTTAGCTT
     401  CTACAGAAAA TAGTAGCAGT GTCATTGAAA AAGAAATTTT TGAAAGCATC
     25  451  AAAAAGATTA ATGAAGAAGG CAAAGCTTTG CTTGAACAAC GGACAGAGTT
     501  AAAGCATGCG ACGAATCCCG AACTACTCAG CATCTATGAG CGTCTATTAA
     551  ACAATAAAAA AGATCGCGTT GTTGTTCCTA TTGAAAATCG TGTCTGCAGT
     601  GGTGTGCATA TTGTTCTAAC TCCTCAACAC GAAAATCTTG TAAGAAAGAA
     651  AGACCGACTC ATTTTTTTCG AACATTGCTC TCGAATTCTC TATTGGCAAG
     30  701  AATCCCAAGT CAATGCTCAG GAAAATTCCA CAGCAAAACG TCGTCGTCGT
     751  CGCGCAGCTG TATAA
```

The PSORT algorithm predicts an inner membrane location (0.070).

- 35 The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 59A) or his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used in Western blot (Figure 59B) and FACS (Figure 59C) analyses.

These experiments show that cp6814 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 60

The following *C.pneumoniae* protein (PID 4376830) was expressed <SEQ ID 119; cp6830>:

```

40      1  MKWLPATAVF AAVLPALTAF GDPASVEIST SHTGSGDPTS DAALTGFTQS
      51  STETDGTTYT IVGDITFSTF TNIPVPVVT DANDSSSNSS KGGSSSSGAT
     101  SLIRSSNLHS DFDFTKDSVL DLYHLFFPSA SNTLNPALLS SSSSGGSSSS
     151  SSSSSSGSAS AVVAADPKGG AAFYSNEANG TLFTTDSGN PGSLLQLNLK
     201  MTGDGAAIYS KGPLVFTGLK NLFTGNESQ KSGGAAYTEG ALTTQAIVEA
     45  251  VTFTGNTSAG QGGAIYVKEA TLFNALDSLK FEKNTSGQAG GGIYTESTLT
     301  ISNITKSIEF ISNKASVPAP APEPTSPAPS SLINSTTIDT STLQTRAASA
     351  TPAVAPVAHV TPTPISTQET AGNGGAIYAK QGISISTFKD LTFKSNSASV
```



851 GFFVSDVYRN NPQSTATLVM SPDSWKIRGG NLSRQAFLLR GSNNYVYNSN  
 901 CELFGHYAME LRGSSRNYNV DVGTKLRF\*

A predicted signal peptide is highlighted.

The cp6733 nucleotide sequence <SEQ ID 116> is:

```

5      1  ATGAAGACTT CGATTCCTTG GGTTTTAGTT TCCTCCGTGT TAGCTTTCTC
      51  ATGTCACCTA CAGTCACTAG CTAACGAGGA ACTTTTATCA CCTGATGATA
     101  GCTTTAATGG AAATATCGAT TCAGGAACGT TTAACGAGG GCTGAAAGG
     151  ACAACATATT CTCTAACAGG AGATGTCCTT TTTACGAGC CTGGAAGAGG
     201  CACTCCCTTA TCTGACAGTT GTTTTAAGCA AACCACGGAC AATCTTACCT
     251  TCTTGGGGAA CGGTCATAGC TTAACGTTTG GCTTTATAGA TGCTGGCACT
     301  CATGCAGGTG CTGCTGCATC TACAACAGCA AATAAGAATC TTACCTTCTC
     351  AGGGTTTTCC TTAAGAGTTT TTGATTCCTC TCCTAGCACA ACGGTACTTA
     401  CAGGTCAGGG AACGCTTTCC TCAGCAGGAG GCGTAAATTT AGAAAAATAT
     451  CGTAAACTTG TAGTTGCTGG GAATTTTCTT ACTGCAGATG GTGGAGCTAT
     501  CAAAGGAGCG TCTTTCCTTT TAAGTGGCAC TCTGGAGATG GCTCTTTTAA
     551  GTAACAACCTC TTAACAACA AAGGAGGAG CAATTGCTAC TACAGCAGGC
     601  GCTCGCATAG CAAATAACAC AGGTTATGTT AGATTCTTAT CTAACATAGC
     651  GTCTACGTCA GGAGGCGCTA TCGATGATGA AGGCACGTCG ATACTATCGA
     701  ACAACAAATT TCTATATTTT GAAGGGAATG CAGCGAAAAC TACTGGCGGT
     751  GCGATCTGCA ACACCAAGGC GAGTGGATCT CCTGAACTGA TAATCTCTAA
     801  CAATAAGACT CTGATCTTTG CTTCAAACGT AGCAGAAACA AGCGGTGGCG
     851  CCATCCATGC TAAAAAGCTA GCCCTTTCCT CTGGAGGCTT TACAGAGTTT
     901  CTACGAAATA ATGTCTCATC AGCAACTCCT AAGGGGGGTG CTATCAGCAT
     951  CGATGCCTCA GGAGAGCTCA GTCTTTCTGC AGAGACAGGA AACATTACCT
    1001  TTGTAAGAAA TACCCTTACA ACAACCGGAA GTACCGATAC TCCTAAACGT
    1051  AATGCGATCA ACATAGGAAG TAACGGGAAA TTCACGGAAT TACGGGCTGC
    1101  TAAAAATCAT ACAATTTTCT TCTATGATCC CATCACTTCA GAAGGAACCT
    1151  CATCAGACGT ATTGAAGATA AATAACGGCT CTGCGGGAGC TCTCAATCCA
    1201  TATCAAGGAA CGATTCATTT TTCTGGAGAA ACCCTAACAG CAGATGAACCT
    1251  TAAAGTTGCT GACAATTTAA AATCTTCATT CACGCAGCCA GTCTCCCTAT
    1301  CCGGAGGAAA GTTATTGCTA CAAAAGGGAG TCACTTTAGA GAGCAGGAGC
    1351  TTCTCTCAAG AGGCCGGTTC TCTCCTCGGC ATGGATTGAG GAACGACATT
    1401  ATCAACTACA GCTGGGAGTA TTACAATCAC GAACCTAGGA ATCAATGTTG
    1451  ACTCCTTAGG TCTTAAGCAG CCCGTCAGCC TAACAGCAA AGGTGCTTCA
    1501  AATAAAGTGA TCGTATCTGG GAAGCTCAAC CTGATTGATA TTGAAGGGAA
    1551  CATTTATGAA AGTCATATGT TCAGCCATGA CCAGCTCTTC TCTCTATTAA
    1601  AAATCACGGT TGATGCTGAT GTTGATACTA ACGTTGACAT CAGCAGCCTT
    1651  ATCCCTGTTT CTGCTGAGGA TCCTAATTCA GAATACGGAT TCCAAGGACA
    1701  ATGGAATGTT AATTGGACTA CGGATACAGC TACAAATACA AAAGAGGCCA
    1751  CGGCAACTTG GACCAAAACA GGATTTGTTT CCAGCCCCGA AAGAAAATCT
    1801  GCGTTAGTAT GCAATACCCT ATGGGGAGTC TTTACTGACA TTCGCTCTCT
    1851  GCAACAGCTT GTAGAGATCG GCGCAACTGG TATGGAACAC AAACAAGGTT
    1901  TCTGGGTTTC CTCCATGACG AACTTCCTGC ATAAGACTGG AGATGAAAAT
    1951  CGCAAAGGCT TCCGTCATAC CTCTGGAGGC TACGTCATCG GTGGAAGTGC
    2001  TCACACTCCT AAAGACGACC TATTTACCTT TCGGTTCTGC CATCTCTTTG
    2051  CTAGAGACAA AGATTGTTTT ATCGCTCACA ACAACTCTAG AACCTACGGT
    2101  GGAACCTTAT TCTTCAAGCA CTCTCATACC CTACAACCCC AAAACTATTT
    2151  GAGATTAGGA AGAGCAAAGT TTTCTGAATC AGCTATAGAA AAATTCCTTA
    2201  GGGAAATTCC CTTAGCCTTG GATGTCCAAG TTTCGTTTCA CCATTCAGAC
    2251  AACCGTATGG AAACGCACTA TACCTCATTG CCAGAATCCG AAGGTTCTTG
    2301  GAGCAACGAG TGTATAGCTG GTGGTATCGG CCTAGACCTT CCTTTTGTTC
    2351  TTTCCAACCC ACATCCTCTT TTCAAGACCT TCATTCCACA GATGAAAGTC
    2401  GAAATGGTTT ATGTATCACA AAATAGCTTC TCGAAAGCTT CTAGTGATGG
    2451  CCGTGGTTTT AGTATTGGAA GGCTGCTTAA CCTCTCGATT CCTGTGGGTG
    2501  CGAAATTTCG GCAGGGGGAT ATCGGAGATT CCTACACCTA TGATCTCTCA
    2551  GGATTCCTTG TTTCCGATGT CTATCGTAAC AATCCCCAAT CTACAGCGAC
    2601  TCTTGTGATG AGCCCAGACT CTTGGAAAAT TCGCGGTGGC AATCTTTCAA
    2651  GACAGGCATT TTTACTGAGG GGTAACAACA ACTACGCTTA CAACTCCAAT
    2701  TGTGAGCTCT TCGGACATTA CGCTATGGAA CTCCGTGGAT CTTCAAGGAA
    2751  CTACAATGTA GATGTTGGTA CCAAACCTCG ATTCTAG
  
```

The PSORT algorithm predicts an outer membrane location (0.924).

5

10

15

20

25

30

35

40

45

50

55

60

65

1801 TACGTTACTA AAACCTTCCA GTGTTCCGAT TCTCATCGCC TCCAGTTTAC  
 1851 TAGTAATAAAA GCAGCAGATG AAGGCGGGGG CCTGTATTGT GGTGACGATG  
 1901 TCACGCTAAC GAACCTGACA GGGAAAACAC TATTTCAAGA GAATAGCAGT  
 1951 GAGAAACATG GAGGTGGGCT CTCTCTCGCC TCAGGAAAAT CTCTGACTAT  
 2001 GACATCGTTA GAGAGCTTCT GCTTAAATGC AAATACAGCA AAGGAAAACG  
 2051 GAGGCGGTGC GAATGTCCCT GAAAATATTG TACTCACCTT CACCTATACT  
 2101 CCCACTCCAA ATGAACCTGC GCCTGTGCAG CAGCCCGTGT ATGGAGAAGC  
 2151 TCTTGTTACT GGAAATACAG CCACAAAAAG TGGTGGGGGC ATTTACACGA  
 2201 AAAATGCGGC CTTCTCAAAT TTATCTTCTG TAACTTTGA TCAAAATACC  
 2251 TCTTCAGAAA ATGGTGGTGC CTTACTTACC CAAAAAGCTG CAGATAAAAC  
 2301 GGACTGTTCT TTCACCTATA TTACAAATGT CAATATCACC AACAAATACAG  
 2351 CTACAGGAAA TGGTGGGGGC ATTGCTGGGG GAAAAGCACA TTTCCGATCGC  
 2401 ATTGATAATC TTACAGTCCA AAGCAACCAA GCAAAGAAAG GTGGTGGGGT  
 2451 TTATCTTGAA ATGACCCTCA TCCTGGAAAA GGTATTATACA GGTCTGTCT  
 2501 CACAAAATAC AGCTACAGAA AGTGGTGGGG GTATCTACGC TAAGGATATT  
 2551 CAACTACAAG CTCTACCTGG AAGCTTCACA ATTACCGATA ATAAAGTCGA  
 2601 AACTAGTCTT ACTACTAGCA CTAATTTATA TGGTGGGGGC ATCTATTCCA  
 2651 GTGGAGCTGT CACGCTAACC AATATATCTG GAACCTTTGG CATTACAGGA  
 2701 AACTCTGTTA TCAATACAGC GACATCCCAG GATGCAGATA TACAAGGTGG  
 2751 GGGCAATTTAT GCAACCACGT CTCTCTCAAT AAATCAATGT AATACACCCA  
 2801 TTCTATTTAG CAACAACCTC GCTGCCACTA AAAAAACATC AACAAACAAAG  
 2851 CAAATTGCTG GTGGGGCTAT CTTCTCCGCT GCAGTAACTA TCGAGAATAA  
 2901 CTCTCAGCCC ATTATTTTCT TAAATAATTC CGCAAAGTCG GAAGCAACTA  
 2951 CAGCAGCAAC TGCAGGAAAT AAAGATAGCT GTGGAGGAGC CATTGACGCT  
 3001 AACTCTGTTA CTTTAAACAAA TAACCTTGAA ATAACCTTTA AAGGAAATTA  
 3051 TGCAGAAACT GGAGGAGCGA TTGGCTGTAT TGATCTTACT AATGGCTCAC  
 3101 CTCCCCGTAA AGTCTCTATT GCAGACAACG GTTCTGTCTT TTTTCAAGAC  
 3151 AACTCTGCGT TAAATCGCGG AGGCGCTATC TATGGAGAGA CTATCGATAT  
 3201 CTCCAGGACA GGTGCGACTT TCATCGGTAA CTCTTCAAAA CATGATGGAA  
 3251 GTGCAATTTG CTGTTCAACA GCCCTAACTC TTGCGCCAAA CTCCCAACTT  
 3301 ATCTTTGAAA ACAATAAGGT TACGGAAACC ACAGCCACTA CAAAAGCTTC  
 3351 CATAAATAAT TTAGGAGCTG CAATTTATGG AAATAATGAG ACTAGTGACG  
 3401 TCACTATCTC TTTATCAGCT GAGAATGGAA GTATTTTCTT TAAAACAAT  
 3451 CTATGCACAG CAACAAACAA ATACTGCAGT ATTGCTGGAA ACGTAAATTT  
 3501 TACAGCAATA GAAGCTTCAG CAGGGAAAGC TATATCTTTC TATGATGCAG  
 3551 TTAACGTTTC CACCAAAGAA ACAAATGCTC AAGAGCTAAA ATTAATGAA  
 3601 AAAGCGACAA GTACAGGAAC GATTCTATTT TCTGGGGAAC TTCACGAAAA  
 3651 TAAATCCTAT ATTCCACAGA AAGTCACTTT CGCACATGGG AATCTCATTC  
 3701 TAGGTAAAAA TGCAGAACTT AGCGTAGTTT CCTTTACCCA ATCTCCAGGC  
 3751 ACCACAATCA CTATGGGCCC AGGATCGGTT CTTTCCAACC ATAGCAAAGA  
 3801 AGCAGGAGGA ATCGCTATAA ACAATGTCAT CATTGATTTT AGTGAATCG  
 3851 TTCTACTTAA AGATAATGCA ACAGTAGCTC CACCCACTCT TAAATTAGTA  
 3901 TCGAGAACTA ATGCAGATAG TAAAGATAAG ATTGATATTA CAGGAACGTG  
 3951 GACTCTTCTA GATCCTAATG GCAACTTATA TCAAAATTCT TATCTTGGTG  
 4001 AAGACCGCGA TATCACTCTT TTCAATATAG ACAATTCTGC AAGTGGGGCA  
 4051 GTTACAGCCA CGAATGTCAC CCTTCAAGGG AATTTAGGAG CTA AAAAAGG  
 4101 ATATTTAGGA ACCTGGAATT TGGATCCAAA TTCCTCGGGT TCAAAAATTA  
 4151 TTCTAAAAATG GACCTTTGAC AAATACCTGC GCTGGCCCTA CATCCCTAGA  
 4201 GACAACCACT TCTACATCAA CTCTATTTGG GGAGCACAAA ACTCTTTAGT  
 4251 GACTGTGAAA CAAGGGATCT TAGGGAACAT GTTGAACAAT GCAAGGTTTG  
 4301 AAGATCCTGC TTTCAACAAC TTCTGGGCTT CGGCTATAGG ATCTTTCCCTT  
 4351 AGGAAAGAAG TATCTCGAAA TTCTGACTCA TTCACCTATC ATGGCAGAGG  
 4401 CTATACCGCT GCTGTGGATG CCAAACCTCG CCAAGAATTT ATTTTAGGAG  
 4451 CTGCCTTCAG TCAGGTTTTT GGTACGCGG AGTCTGAATA TCACCTTGAC  
 4501 AACTATAAGC ATAAAGGCTC AGGTCACTCT ACACAAGCAT CTCTTTATGC  
 4551 TGGCAATATC TTCTATTTTC CTGCGATACG GTCTCGGCCT ATTCTATTCC  
 4601 AAGGTGTGGC GACCTATGGT TATATGCAAC ATGACACCAC AACCTACTAT  
 4651 CCTTCTATTG AAGAAAAAAA TATGGCAAAC TGGGATAGCA TTGCTTGGTT  
 4701 ATTTGATCTG CGTTTCAGTG TGGATCTTAA AGAACCTCAA CCTCACTCTA  
 4751 CAGCAAGGCT TACCTTCTAT ACAGAAGCTG AGTATACCAG AATTGCCAG  
 4801 GAGAAATTCA CAGAGCTAGA CTATGATCCT AGATCTTTCT CTGCATGCTC  
 4851 TTATGGAAAC TTAGCAATTC CTAAGGATT CTCTGTAGAC GGAGCATTAG  
 4901 CTTGGCGTGA GATTATTCTA TATAATAAAG TATCAGCTGC GTACCTCCCT  
 4951 GTGATTCTCA GGAATAATCC AAAAGCGACC TATGAAGTTC TCTCTACAAA  
 5001 AGAAAAGGGC AACGTAGTCA ACGTTCTCCC TACAAGAAAC GCAGCTCGTG  
 5051 CAGAGGTGAG CTCTCAAAT TATCTTGGAA GTTACTGGAC ACTCTACGGC  
 5101 ACGTATACTA TTGATGCTTC AATGAATACT TTAGTGCAAA TGGCCAACGG  
 5151 AGGGATCCGG TTTGTATTCT AG

-99-

5 401 DATLTVDSST IGESGGAIFA ADSIQIQQCT GTTLFSGNTA NKSGGGIYAV  
 451 GQVTLLEDIAN LKMTNNTCKG EGGAIYTKKA LTINNGAILT TFSGNTSTDN  
 501 GGAIFAVGGI TSLDLVEVRF SKNKTGNYS A PITKAASNTA PVVSSSTTAA  
 551 SPAVPAAAAA PVTNAAKGGA LYSTEGTLVS GITSILSFEN NECQNQGGGA  
 601 YVTKTFQCS D SHRLQFTSNK AADGGGLYC GDDVTLTNLT GKTFLQENSS  
 651 EKHGGGLSLA SGKSLTMTSL ESFCLNANTA KENGGGANVP ENIVLTFITYT  
 701 PTPNEPAPVQ QPVYGEALVT GNTATKSGGG IYTKNAAFSN LSSVTFDQNT  
 751 SSENGGALLT QKAADKTDCS FTYITNVNIT NNTATGNGGG IAGGKAHFDR  
 801 IDNLTVQSNQ AKKGGGVYLE DALILEKVIT GSVSQNTATE SGGGIYAKDI  
 10 851 QLQALPGSFT ITDNKVETSL TTSTNLYGGG IYSSGAVTLT NISGTFGITG  
 901 NSVINTATSQ DADIQGGGIY ATTSL SINQC NTPILFSNNS AATKKTSTTK  
 951 QIAGGAIFSA AVTIENNSQP IIFLNNSAKS EATTAATAGN KDSCGGAIAA  
 1001 NSVTLTNPE ITFKGNYAET GGAIGCIDLT NGSPPRK VSI ADNGSVLFQD  
 1051 NSALNRGGAI YGETIDISRT GATFIGNSSK HDGSAICCS T ALTAPNSQL  
 15 1101 IFENNKVTET TATTKASINN LGAAIYGNNE TSDVTISLSA ENGSIFFKNN  
 1151 LCTATNKYCS IAGNVKFTAI EASAGKAISF YDAVNVSTKE TNAQELKLE  
 1201 KATSTGTILF SGELHENKSY IPQKVTFAHG NLILGKNAEL SVVSFTQSPG  
 1251 TTITMGPGSV LSNHSKEAGG IAINNVIIDF SEIVPTKDNA TVAPPTLKL  
 1301 SRTNADSKDK IDITGTVTL L DPNGNLYQNS YLGEDRDITL FNIDNSASGA  
 20 1351 VTATNLTLQ NLGAKKGYLG TWNLDPNSSG SKIILKWTFD KYLRWPYIPR  
 1401 DNHFYINSIW GAQNSLVTVK QGILGNMLNN ARFEDPAFNN FWASAIGSFL  
 1451 RKEVSRNSDS FTYHGRGYTA AVDAKPRQEF ILGAAFSQVF GHAESEYHLD  
 1501 NYKHKSGSHS TQASLYAGNI FYFPAIRSRP ILFQGVATYG YMQHDTTTTY  
 25 1551 PSIEEKNMAN WDSIAWLFDL RFSVDLKEPQ PHSTARLT FY TEAEYTRIRQ  
 1601 EKFTELDYDP RSFSACSYGN LAIPTGFSD GALAWREIIL YNKVSAAYLP  
 1651 VILRNNPKAT YEVLSTKEKG NVNVNLPTRN AARAEVSSQI YLGSYWTLYG  
 1701 TYTIDASMNT LVQMANGGIR FVF\*

A predicted signal peptide is highlighted.

The cp6830 nucleotide sequence <SEQ ID 120> is:

30 1 ATGAAGTGGC TACCAGCTAC AGCTGTTTTT GCTGCCGTAC TCCCCGCACT  
 51 AACAGCCTTC GGAGATCCCG CGTCTGTTGA AATAAGTACC AGCCATACAG  
 101 GATCCGGGGA TCCTACAAGC GACGCTGCCT TAACAGGATT TACACAAAGT  
 151 TCCACAGAAA CTGACGGTAC TACCTATACC ATTGTCGGTG ATATCACCTT  
 201 CTCTACTTTT ACGAATATTC CTGTTCCCGT AGTAACTCCA GACGCCAACG  
 35 251 ATAGTTCCAG CAATAGCTCT AAAGGAGGAA GTAGCAGTAG TGGAGCTACA  
 301 TCTCTAATCC GATCCTCAA CCTACACTCC GATTTTGATT TTACAAAAGA  
 351 TAGCGTGT A GACCTCTATC ACCTTTTCTT TCCTTCAGCT TCAAATACCT  
 401 TCAATCCTGC ACTCCTTTCT TCCAGTAGCA GCGGTGGATC CTCGAGCAGC  
 451 AGTAGCTCCT CATCATCTGG AAGTGCATCT GCTGTTGTTG CTGCGGACCC  
 40 501 AAAAGGAGGC GCTGCCTTT ATAGTAACGA GGCTAACGGA ACTTTAACCT  
 551 TCACTACAGA CTCTGGAAT CCCGGCTCCC TGACTCTTCA GAATCTTAAA  
 601 ATGACCGGAG ATGGAGCCGC CATCTACTCG AAGGGTCCCT TAGTATTTAC  
 651 TGGTTTAAAA AATCTAACCT TTACAGGAAA TGAATCTCAG AAATCTGGAG  
 701 GTGCTGCCTA TACTGAAGC GCACTCACA CACAAGCAAT CGTTGAAGCC  
 45 751 GTAACCTTTA CTGGCAACAC CTCGGCAGGG CAAGGAGGCG CTATCTATGT  
 801 TAAAGAAGCT ACCCTATTCA ATGCTCTAGA CAGCCTCAA TTTGAAAAA  
 851 ACACTTCTGG GCAAGCTGGT GGTGGAATCT ATACAGAGTC TACGCTCACA  
 901 ATCTCGAACA TCACAAAATC TATTGAATTT ATCTCTAATA AAGCTTCTGT  
 951 CCCTGCCCCC GCTCCTGAGC CCACCTCTCC GGCTCCAAGT AGCTTAATAA  
 50 1001 ATTCTACAAC GATCGATACC TCGACTCTCC AAACCCGAGC AGCATCCGCA  
 1051 ACTCCAGCAG TGGCTCCTGT TGCTGCCGTA ACTCCAACAC CAATCTCTAC  
 1101 TCAAGAGACC GCAGGAAATG GAGGCGCTAT CTATGCTAAA CAAGGTATTT  
 1151 CGATATCCAC GTTTAAAGAT CTGACCTTCA AGTCTAACTC TGCATCGGTA  
 1201 GATGCCACCC TTACTGTCTGA TTCTAGCACT ATTGGAGAAT CTGGAGGTGC  
 55 1251 TATCTTTGCA GCAGACTCTA TACAAATCCA ACAGTGCACG GGAACCACCT  
 1301 TATTCAGTGG CAATACTGCC AATAAGTCTG GTGGGGGTAT TTACGCTGTA  
 1351 GGACAAGTCA CCCTAGAAGA TATAGCGAAT CTGAAGATGA CCAACAACAC  
 1401 CTGTAAAGGT GAAGGTGGAG CCATCTACAC TAAAAAGGCT TTAACATATCA  
 1451 ACAACGGTGC CATCTCTACT ACATTTTCTG GAAATACATC GACAGATAAT  
 60 1501 GGTGGGGCTA TTTTGTGCTG AGGTGGCATC ACTCTCTCTG ATCTTGTAGA  
 1551 AGTCCGCTTT AGTAAAAATA AGACCGGAAA TTATTCCGCT CCTATTACCA  
 1601 AAGCGGCTAG CAACACAGCT CCTGTAGTTT CTAGCTCTAC AACTGCTGCA  
 1651 TCTCTGCGG TCCCTGCTGC CGCTGCAGCA CCTGTTACAA ACGCAGCAA  
 1701 AGGAGGGGCT TTATATAGTA CAGAAGGACT GACTGTATCT GGAATCACAT  
 65 1751 CGATATTGTC GTTTGAAAAAC AACGAATGCC AGAATCAAGG AGGTGGGGCT

## Example 62

The following *C.pneumoniae* protein (PID 4377101) was expressed <SEQ ID 123; cp7101>:

```

1  MYSCYSKGIS HNYLLHPMSR LDIFVFDLSI ANQDQNLLEE IFCSEDTVLF
5  51  KAYRTTALQS PLAAKNLNIA RKVANYILAD NGEIDTVKLV EAIHHL SQCT
101 YPLGPHRHNE AQDREHLLKM LKALKENPKL KESIKTLFVP SYSTIQNLIR
151 HTLALNPQTI LSTIHVRQAA LTALFTYLRQ DVGSCFATAP AILIHQEYPE
201 RFLKDLNDLI SSGKLSRIVN QREIAVPINL SGCIGELFKP LRILDLYPDP
251 LVKLSSSPGL KKAFAANLI ETLGDSEAQI QQLLSHQYLM QKLQNVHETL
301 TANDIIKSTL LHYYQLQEST VRAIFFKEGL FSKEQVAFST QHPRELSEIQ
10 351 RVYHYLHAYE EAKSAFIHDT QNPLLKAWEY TLATLADASQ PTISNHIRLA
401 LGWKSEDPHS LVSLVTHFVE EEEVENIRILV QQCEQTYHEA RSQLEYIEGR
451 MRNPLNNQDS QILTMHMRF RQELNKALYE WDSAQEKAKK FLHLPEFLLS
501 FYTKQIPLYF RSSYDAFIQE FAHLYANAPA GFRILFTHGR THPNTWSPIY
551 SINEFIRFLS EFFTSTESEL LGKHAVINLE KETSRLVHNI TAMLHDTVPFQ
15 601 EALLTRILEA YQLPVPPSIL NHLDQLSQTP WVVVSGGTVD TLLLDYFESS
651 EPLTLTEKHP ENPHELAIFY ADALKDLPTG IKSYLEEGSH SLLSSSPHIV
701 FSIIAGSPLF REAWDNDWYS YTWLRDVWVK QHQDFLQDTI LPQLSIYAFI
751 ENFCNKYALQ HVVHDFHDFC SDHSLTLPEL YDKGSRFLSS LFTKDKTVAL
801 IYIRLLLYLM VREVPYVSEQ QLPEVLDNVS SYLGISSRIT YEKFRSLIEE
20 851 TIPKMTLLSS ADLRHIYKGL LMQSYQKIYT BEDTYLRLIT AMRHHNLAYP
901 APLLFADSNW PSYIFGFILN PGTTEIDLWK FNYAGLQGP LDNIQELFAT
951 SRPWTLYANP IDYGMPPPPG YRSRLPKEFF *
```

The cp7101 nucleotide sequence <SEQ ID 124> is:

```

1  ATGTATTCGT GTTACAGCAA AGGAATATCC CATAACTATC TTCTACATCC
25 51  TATGTCACGT TTGGATATTT TTGTTTTCGA TTCTCTGATC GCAAACCAGG
101 ATCAAAATCT TCTTGAGGAA ATTTTCTGTT CTGAAGACAC AGTTTTATTT
151 AAAGCCTACC GTACTACGGC TCTACAATCC CCTCTAGCTG CTAAGAACCT
201 AAATATCGCC CGTAAAGTCG CAAATTATAT CTAGCTGAC AATGGGGAAA
251 TCGATACAGT AAAGCTTGTC GAAGCCATTC ACCATCTCTC ACAATGTACC
30 301 TATCCTTTAG GGCCTCATCG CCATAATGAA GCTCAAGATC GTGAACACCT
351 CCTTAAATG CTAAGAGCTC TAAAGGAAAA TCCTAAATTA AAAGAAAGCA
401 TCAAAACTCT CTTTGTCCCT TCATACTCTA CAATCCAAAA CCTAATTCGC
451 CATACACTAG CATTGAATCC ACAGACAATT CTCTCTACGA TTCATGTGCG
501 TCAAGCAGCA CTCACAGCGC TCTTCACCTA CCTTCGGCAA GATGTAGGTT
35 551 CCTGTTTTGC TACGGCTCCT GCCATTCTCA TTCACCAAGA ATATCCAGAA
601 CGATTCCCTA AAGATCTCAA TGATCTCATT AGCAGTGGCA AACTCTCTAG
651 AATCGTAAAC CAAAGGGAAA TTGCGGTTCC TATAAACCTT TCGGGATGCA
701 TTGGAGAGCT ATTCAAGCCT TTAAGGATTC TAGATCTTTA TCCTGATCCT
751 CTGTTAAGC TCTCCTCATC TCCAGGACTC AAAAAAGCCT TTCTGTCTGC
40 801 CAATCTTATT GAAACTCTTG GGGATTCTGA AGCACAATC CAACAGTTGC
851 TCTCGCATCA ATATTTGATG CAAAACTAC AAAATGTCCA TGAGACCTTA
901 ACTGCTAACG ACATTATCAA ATCGACACTT CTGCACTACT ATCAGCTCCA
951 AGAAAGTACT GTACGAGCTA TTTTCTTCAA AGAAGGGTTG TTCAGCAAAG
1001 AACAAAGTGGC ATTCTCGACG CAACACCCCA GAGAGCTCTC AGAAATACAA
45 1051 CGGGTATACC ACTACTTACA TGCCATATGAA GAAGCAAAAT CTGCTTTTAT
1101 CCATGACACT CAAATCCCT TACTGAAAGC CTGGGAGTAT ACTTTAGCGA
1151 CTCTTGCGGA TGCTAGCCAA CCTACCATCT CAAACCATAT CCGCCTTGCC
1201 TTAGGATGGA AAAGTGAAGA CCCTCACAGT CTTGTATCTC TAGTTACACA
1251 CTTTGTGAA GAGGAAGTAG AAAACATCCG AATTTTAGTC CAACAATGTG
50 1301 AACAGACCTA TCACGAAGCA CGCTCCCAAC TAGAATATAT TGAAGGGCGG
1351 ATGCGCAACC CACTAAATAA TCAAGACAGT CAGATTTTGA CGATGGATCA
1401 CATGCGCTTC CGTCAAGAAC TCAATAAAGC TCTTTATGAG TGGGATAGTG
1451 CTCAAGAAAA GGCAAGAAA TTTCTACATC TTCCTGAATT CTTACTTTCT
1501 TTCTATACAA AGCAAATTCC CTTATACTTT CGTAGTTCTT ACGATGCCTT
55 1551 CATTCAAGAA TTTGCTCATC TCTATGCTAA TGCTCCCGCT GGCTTCCGTA
1601 TTTTTCAC CACTGGACGC ACCCATCCGA ACACATGGTC CCCCATCTAT
1651 TCGATTAATG AATTTATACG TTTTCTTTCT GAATTCCTCA CCTCCACAGA
1701 GTCAGAACTT CTGGGGAAC ATGCCGTGAT CAATTTAGAG AAAGAAACAT
1751 CTCGGCTCGT CCACAACATC ACTGCCATGC TACACACGGA TGTTTTCCAA
60 1801 GAAGCTCTCC TTACAAGAA TTTAGAAGCC TATCAGCTTC CTGTGCCTCC
1851 CTCCATCTTA AACCATTAG ATCAGCTGTC ACAAACTCCC TGGGTTTATG
1901 TTTCTGGAGG AACAGTGAC ACTCTTCTTT TGGATTATTT TGAAGCTCA
1951 GAACCTCTGA CACTTACAGA AAAGCATCCT GAAAATCTC ATGAGCTTGC
2001 AGCTTTCTAC GCAGACGCCC TTAAAGATCT CCCTACAGGA ATTAAGGTT
```

The PSORT algorithm predicts an outer membrane location (0.926).

The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 60A) or his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used in Western blot (Figure 60B) and FACS (Figure 60C) analyses.

- 5 The cp6830 protein was also identified in the 2D-PAGE experiment (Cpn0540) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6830 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 61

- 10 The following *C.pneumoniae* protein (PID 4376854) was expressed <SEQ ID 121; cp6854>:

```

1  MSIAIAREQY AAILDMHPKP SIAMFSSEQA RTSWEKRQAH PYLYRLLEII
51  WGVVKFLLGL IFFIPLGLFW VLQKICQNF I LLGAGGWIFR PICRDSNLLR
101 QAYAARLFSA SFQDHVSSVR RVCLQYDEVF IDGLELRLPN AKPDRWMLIS
151 NGNSDCLEYR TVLQGEKDWI FRIAEESSQN ILIFNYPGVM KSQGNITRNN
15  201 VVKSYQACVR YLRDEPAGPQ ARQIVAYGYS LGASVQAEAL SKEIADGSDS
251 VRWFVVKDRG ARSTGAVAKQ FIGSLGVWLA NLTHWNINSE KRSKDLHCPE
301 LFIYGKDSQG NLIGDGLFKK ETCFAAPFLD PKNLEECGSK KIPVAQTGLR
351 HDHILSDDVI KEVAGHIQRH FDN*
```

The cp6854 nucleotide sequence <SEQ ID 122> is:

```

20 1  ATGTCAATAG CTATTGCAAG GGAACAATAC GCAGCTATAT TGGATATGCA
51  TCCTAAACCT TCGATCGCCA TGTTTTCTTC GGAGCAGGCG AGAACTTCTT
101 GGGAGAAACG ACAGGCTCAT CCTTACCCTT ATCGTCTTCT TGAGATCATA
151 TGGGGTGTG TGAAATTTCT TCTCGGCTTA ATCTTCTTTA TTCCCTTGGG
201 TCTTTTCTGG GTCCTTCAGA AGATATGTCA GAATTTTATT CTTCTTGGTG
25 251 CAGGAGGGTG GATTTTTAGA CCCATATGCA GGGACTCTAA TTTATTGCGA
301 CAAGCTTACG CCGCGCGTCT TTTCTCCGCT TCATTCCAAG ATCATGTCTC
351 CTCTGTGCGA AGGGTTTGCT TACAGTATGA CGAGGTCTTT ATTGACGGAT
401 TGGAGTTACG TCTTCCCAAT GCTAAGCCAG ATCGATGGAT GTTAATCTCC
451 AATGGAAACT CCGATTGCTT AGAGTATAGG ACAGTCTGTC AAGGGGAAAA
30 501 GGAAGTGGATA TTCCGTATTG CTGAAGAGTC TCAATCCAAC ATTTTAATCT
551 TCAATTACCC AGGAGTCATG AAGAGCCAAG GGAATATAAC AAGAAACAAT
601 GTAGTCAAAT CTTATCAAGC ATGCGTACGC TATCTTAGAG ATGAACCCGC
651 AGGACCTCAG GCGCGTCAAA TCGTTGCTTA TGGCTATTCT TTAGGAGCTA
701 GTGTTCAAGC CGAAGCATT AAGTAAAGAGA TCGCAGACGG AAGTGATAGC
35 751 TCCCGTTGGT TTGTCGTTAA AGATCGAGGA GCTCGCTCTA CAGGAGCCGT
801 TGCTAAACAG TTTATTGGAA GTCTAGGAGT TTGGCTGGCG AATCTTACCC
851 ATTGGAATAT TAATTCTGAA AAGAGAAGCA AGGACTTGCA TTGCCAGAA
901 CTCTTTATTT ATGGCAAGGA TTCCAAGGT AATCTTATCG GGGATGGATT
951 GTTCAAAAAA GAGACGTGCT TCGCAGCACC ATTTTATAGT CCTAAAAACT
40 1001 TGGAAGAGTG TTCAGGGAAG AAAATCCCTG TAGCTCAGAC CGGTCTAAGA
1051 CACGATCATA TCCTTCCGA TGATGTGATT AAAGAAGTTG CAGGTCATAT
1101 TCAAAGACAT TTCGATAATT A
```

The PSORT algorithm predicts an inner membrane location (0.461).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 61A.

- 45 The recombinant protein was used to immunise mice, whose sera were used in Western blot (Figure 61B) and FACS (Figure 61C) analyses. A his-tagged protein was also expressed.

These experiments show that cp6854 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

5 651 TTCTATAATT GTAGGAACCA TGGTAGACGT GTCATGGAGA AATACCGCAG  
 701 TACAATGGAT CGGGGATCAG CTCTCTGTTA TTGGGACTTT AGGAGGAACT  
 751 ACTTCTGTGT CTAGTGCAAT CTCAACAGAT GGCACGTGTA TTGTAGGAGG  
 801 TTCTGAAAAT GCAGATTCTC AGACTCATGC CTATGCTTAT AAAACGGTGT  
 851 TTATGAGCGA TATAGGGACC CTCGGAGGTT TTTATTCTTT AGCACATGCA  
 901 GTATCTTCAG ATGGTTCTGT GATTGTAGGA GTATCCACGA ACTCTGAGCA  
 951 TAGATATCAT GCATTCCAAT ATGCTGATGG ACAGATGGTA GATTTAGGAA  
 1001 CTTTAGGAGG GCCTGAATCT TATGCTCAAG GTGTGTCTGG AGATGGAAAG  
 10 1051 GTAATTGTGG GTAGAGCACA AGTACCATCT GGAGATTGGC ATGCGTTCCT  
 1101 ATGTCCTTTC CAAGCTCCGA GCCCTGCTCC TGTCCATGGG GGAAGCACTG  
 1151 TCGTAACTAG CCAGAATCCA CGTGGAAATGG TAGATATCAA TGCTACGTAC  
 1201 TCCTCTTTGA AAAATAGCCA ACAACAATA CAAAGATTGC TTATCCAGCA  
 1251 TAGTGCAAAA GTTGAAAGTG TATCCTCAGG AGCACCATCT TTTACAAGTG  
 1301 TGAAAGGTGC GATCTCAAAA CAGAGCCCTG CAGTGCAAAA TGATGTACAG  
 15 1351 AAAGGGACGT TTTTAAGTTA CCGTTCCCAA GTTCATGGAA ACGTGCAGAA  
 1401 TCAGCAATTG CTCACAGGAG CTTTTATGGA CTGGAACTC GCTTCAGCTC  
 1451 CTAAATGCCG CTTTAAAGTA GCTCTCCACT ATGGCTCTCA AGATGCTCTC  
 1501 GTAGAACGTG CAGCTCTTCC TTACACAGAA CAAGGCTTAG GAAGCAGTGT  
 1551 CTTGTACAGT TTTGGAGGAC AAGTTCAAGG ACGCTATGAC TTTAATTTAG  
 20 1601 GAGAACTGTG TGTTCTGCAA CCCTTTATGG GCATTCAAGT TCTCCACCTA  
 1651 AGTAGAGAAG GGTATTCTGA GAAGAATGTT CGATTTCCTG TAAGCTATGA  
 1701 TTCTGTAGCC TACTCAGCAG CTACTAGCTT TATGGGTGCG CATGTATTTG  
 1751 CCTCCCTAAG CCCTAAATG AGTACAGCAG CAACTTAGG TGTGGAGAGA  
 1801 GATCTGAATT CACATATAGA TGAATTTAAG GGATCCGTCT CTGCTATGGG  
 25 1851 AAACCTTTGTC TTGGAATAAT CTACAGTGAG TGTTTTAAGA CCTTTTGCTT  
 1901 CTCTTGCTAT GTACTATGAC GTAAGACAAC AGCAACTCGT GACGTTGTCA  
 1951 GTAGTTATGA ATCAACAACC CTTAACAGGC AACTAAGCT TAGTAAGCCA  
 2001 AAGTAGCTAT AATCTTAGCT TCTAA

The PSORT algorithm predicts an inner membrane location (0.100).

30 The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 63A) or his-tagged product. The proteins were used to immunise mice, whose sera were used in Western blot (Figure 63B) and FACS (Figure 63C) analyses.

These experiments show that cp7107 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### 35 Example 64

The following *C.pneumoniae* protein (PID 4376467) was expressed <SEQ ID 127; cp6467>:

40 1 MLRFFAVFIS TLWLITSGCS PSQSSKGIFV VNMKEMPRSL DPGKTRLIAD  
 51 QTLMRHLYEG LIVEHSQNGE IKPALAESYT ISEDGTRYTF KIKNILWSNG  
 101 DPLTAQDFVS SWKEILKEDA SSVYLYAFIP IKNARAIFDD TESPENLGV  
 151 ALDKRHLEIQ LETPCAHLFH FLTLPPIFPV HETLRNYSST FEEMPITCGA  
 201 FRPVSLKGL RLHLEKNPMY HNKSRVKLHK IIVQFISNAN TAAILFKHKK  
 251 LDWQGPWGE PIPPEISASL HQDDQLFSLP GASTTWLLFN IQKKPWNNAK  
 301 LRKALSLAID KDMLTKVVYQ GLAEPTDHIL HPRLYPGTYP ERKRQNERIL  
 351 EAQQLFEEAL DELQMTREDL EKETLTFSTF SFSYGRICQM LREQWKKVLK  
 45 401 FTIPIVGQEF FTIQKNFLEG NYSLTVNQWT AAFIDPMSYL MIFANPGGIS  
 451 PYHLQDSHFQ TLLIKITQEH KKHLRNQLII EALDYLEHCH ILEPLCHPNL  
 501 RIALNKNIKN FNLFVVRTSD FRFIEKL\*

A predicted signal peptide is highlighted.

The cp6467 nucleotide sequence <SEQ ID 128> is:

50 1 ATGCTCCGTT TCTTCGCTGT ATTTATATCA ACTCTTTGGC TCATTACCTC  
 51 AGGATGTTCC CCATCCCAAT CCTCTAAAGG AATTTTGTG GTAAATATGA  
 101 AGGAAATGCC ACGCTCCTTG GATCCTGGAA AAACCTGCTC CATTGCAGAC  
 151 CAAACTCTAA TGCGTCATCT ATATGAAGGA CTCGTCGAAG AACATTCCCA  
 201 AAATGGAGAG ATTAAACCAG CCCTTGCGA AAGCTACACC ATCTCCGAAG  
 55 251 ACGGGACTCG GTACACATTT AAAATCAAAA ACATCCTTTG GAGTAACGGA  
 301 GACCCCTGTA CAGCTCAAGA CTTTGTCTCC TCTTGGAAGG AAATCCTAAA

```

2051 ATCTAGAAGA AGGATCCCAC TCTCTACTTA GCTCATCACC CACCCACGTT
2101 TTCTCTATAA TCGCAGGATC TCCTTTATTT CGGGAAGCTT GGGATAATGA
2151 TTGGTACAGC TATACCTGGC TTCGTGATGT CTGGGTGAAA CAACACCAAG
5 2201 ATTTCTTCA AGATACTATA TTACCTCAGC TAAGTATCTA TGCTTTCATA
2251 GAGAAATTTT GTAACAAATA TGCTTTGCAA CATGTAGTTC ATGACTTTCA
2301 TGATTCTGTC TCCGACCACT CCTTGACTCT TCCGGAGCTC TATGACAAAG
2351 GATCGCGTTT TCTAAGCTCC TTATTCACCA AAGATAAGAC CGTAGCTCTT
2401 ATCTATATAC GCCGTCTTCT CTACCTTATG GTCCGTGAAG TCCCTTATGT
10 2451 TTCAGAACAA CAGCTTCCAG AAGTCTTAGA TAACGTCTCT TCATATCTCG
2501 GGATTTCCCTC TCGTATTACC TATGAGAAAT TCCGCTCCCT GATAGAGGAA
2551 ACCATCCCTA AAATGACCTT ACTCTCCTCA GCAGACCTGA GGCATATCTA
2601 TAAAGGTCTC CTCATGCAAA GTTATCAAAA GATCTACACC GAAGAAGATA
2651 CGTACCTCCG CCTCACCACG GCAATGAGGC ATCATAATCT TGCCTATCCC
15 2701 GCTCCTTGGC TCTTTGCAGA CAGTAACTGG CCTTCTATTT ATTTTGGATT
2751 CATCCTAAAT CCAGGAACCA CAGAGATCGA TCTTTGGAAA TTAACTATG
2801 CAGGGCTGCA AGGACAGCCT CTTGACAATA TCCAGGAGCT GTTCGCAACG
2851 TCAAGACCTT GGACCTCTA TGCAAATCCT ATAGATTATG GCATGCCACC
2901 GCCTCCAGGC TACCGCAGCC GCCTCCCTAA AGAATTTTTC TAG

```

The PSORT algorithm predicts a cytoplasmic location (0.206).

20 The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 62A) or his-tagged product. The proteins were used to immunise mice, whose sera were used in Western blot (Figure 62B) and FACS (Figure 62C) analyses.

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

25 These experiments show that cp7101 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 63

The following *C.pneumoniae* protein (PID 4377107) was expressed <SEQ ID 125; cp7107>:

```

30 1 MSIVRNSALP LPCLSRSETF KKVRSHMKFM KVLTPWIYRK DLWVTAFLLT
51 AIPGSFAHTL VDIAGEPRHA AQATGVSGDG KIVIGMKVPD DPFATTVGFG
101 YIDGHLQPLE AVRPQCSVYP NGITPDGTVI VGTNYAIGMG SVAVKWVNGK
151 VSELPMLPDT LDSVASAVSA DGRVIGGNRN INLGASVAVK WEDDVITQLP
201 SLPDAMNACV NGISSDGSII VGTMVDVSWR NTAVQWIGDQ LSVIGTLGGT
35 251 TSVASAISTD GTVIVGGSEN ADSQTHAYAY KNGVMSDIGT LGGFYSLAHA
301 VSSDGSVIVG VSTNSEHRYH AFQYADGQMV DLGTLGGPES YAQGVSGDGK
351 VIVGRAQVPS GDWHAFLCPF QAPSPAPVHG GSTVVTSONP RGMVDINATY
401 SSLKNSQQQL QRLLIQHSK VESVSSGAPS FTSVKGAISK QSPAVQNDVQ
451 KGTFLSYRSQ VHGNVQNQQL LTGAFMWKL ASAPKCGFKV ALHYGSQDAL
501 VERAALPYTE QGLGSSVLSG FGGQVQGRYD FNLGETVVLQ PFMGIQVLHL
40 551 SREGYSEKNV RFPVSYDSVA YSAATSFMGA HVFASLSPKM STAATLGVER
601 DLNSHIDEFK GSVSAMGNFV LENSTVSVLR PFASLAMYD VRQQQLVTLS
651 VVMNQPLTG TSLVLSQSSY NLSF*

```

The cp7107 nucleotide sequence <SEQ ID 126> is:

```

45 1 ATGAGTATAG TCAGAAATTC TGCATTGCCA CTTCCGTGTT TAAGCAGATC
51 CGAAACCTTT AAAAAAGTTA GGTCGCATAT GAAATTTATG AAAGTCTTTA
101 CTCCATGGAT TTATCGAAAA GATCTTTGGG TAACAGCATT CTTACTGACA
151 GCAATTCCAG GATCTTTTGC ACATACTCTT GTTGATATAG CAGGAGAACC
201 TCGGCATGCT GCTCAAGCAA CAGGAGTTTC TGGAGATGGT AAAATTGTTA
50 251 TAGGAATGAA AGTTCCGGAT GATCCTTTTG CTATAACTGT AGGATTTCAA
301 TATATTGATG GGCATTTGCA ACCCTTAGAG GCAGTACGTC CTAATGCTC
351 TGTATACCC T AATGGTATAA CCCCAGGACG AACGGTTATT GTGGGTACAA
401 ACTATGCCAT CGGGATGGGT AGTGTTGCTG TGAAATGGGT AAATGGCAAG
451 GTTTCTGAAC TTCCCATGCT CCCTGACACC CTCGATTCTG TAGCATCGGC
501 AGTTTCTGCA GATGGAAGAG TGATTGGAGG GAATAGAAAT ATAAATCTTG
55 551 GCGCTTCTGT TGCTGTGAAA TGGGAGGACG ACGTGATTAC ACAACTTCCT
601 TCTCTTCTG ATGCTATGAA TGCTTGTGTT AACGGAATTT CTTCAGATGG

```

601 GAAGTTGTTG CCAGAGTTGA GGGCTATGTT TGTGCTAACT ACTCGTAG

The PSORT algorithm predicts an inner membrane location (0.149).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 65A) and as a GST-fusion product (Figure 65B). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 65C) and for FACS analysis.

These experiments show that cp6679 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 66

The following *C.pneumoniae* protein (PID 4376890) was expressed <SEQ ID 131; cp6890>:

```

10      1  MKQLLFCVCV FAMSCSAYAS PRRQDPSVMK ETFRNNYGII VSGQEWVKRG
      51  SDGTITKVLK NGATLHEVYS GLLHGEITL TFPHTTALDV VQIYDQGRV
     101  SRKTFVNGL PSQEELFNEG GTFVLTRWPD NNDSDTITKP YFIETTYQGH
     151  VIEGYSYTSFN GKYSSSIHNG EGVRVVFSSN NILLSEETFN EGVMVKYTTF
     201  YPNRDPESIT HYQNGQPHGL RLTYLQGGIP NTIEEWRYGF QDGTITIVFKN
     251  GCKTSEIAYV KGVKEGLELR YNEQEIVAEV VSWRNDFLHG ERKIYAGGIG
     301  KHEWYYRGRS VSKAKFERLN AAG*

```

A predicted signal peptide is highlighted.

The cp6890 nucleotide sequence <SEQ ID 132> is:

```

20      1  ATGAAACAAT TACTTTTCTG TGTTCGCGTA TTTGCTATGT CATGTTCTGC
      51  TTACGCATCC CCACGACGAC AAGATCCTTC TGTATGAAG GAAACATTCC
     101  GAAATAATTA TGGCATTATT GTTCCGGTC AAGAATGGGT AAAGCGTGGT
     151  TCTGACGGCA CCATCACCAG AGTACTCAA AATGGAGCTA CCCTGCATGA
     201  AGTTTATTCT GGAGGCCTCC TTCATGGGGA AATTACCTTA ACGTTTCCCC
     251  ATACCACAGC ATTGGACGTT GTTCAAATCT ATGATCAAGG TAGACTCGTT
     301  TCTCGCAAAA CCTTTTGTGT GAACGGTCTT CCATCTCAAG AAGAGCTGTT
     351  CAATGAAGAT GGCACGTTTG TCCTCACACG ATGGCCGGAC AACAACGACA
     401  GTGATACCAT CACAAAGCCT TACTTCATAG AAACGACATA TCAAGGGCAT
     451  GTCATAGAAG GAAGTTATAC TTCCTTTAAT GGGAAATACT CCTCATCCAT
     501  CCACAATGGA GAGGGAGTTC GTTCTGTGTT CTCCTCCAAT AACATCCTTC
     551  TTTCTGAAGA GACCTTCAAT GAAGTGTC AAGGTGAAATA TACCACATTC
     601  TATCCGAATC GCGATCCCGA ATCGATTACT CATTATCAAA ATGGACAGCC
     651  TCACGGCTTA CGGCTAACAT ATCTACAAGG TGGCATCCCC AATACGATAG
     701  AGGAGTGGCG TTATGGCTTT CAAGACGGAA CGACCATCGT ATTTAAAAAT
     751  GGTGTGAAGA CATCTGAGAT CGCTTATGTT AAGGGAGTGA AAGAAGGTTT
     801  AGAACTGCGC TACAATGAAC AGGAAATTGT AGCTGAAGAA GTTTCTTGCC
     851  GTAATGATTT TCTGCATGGA GAACGTAAGA TCTATGCTGG AGGAATCCAA
     901  AAGCATGAAT GGTATTACCG CGGGAGATCT GTATCTAAAG CCAAATTCGA
     951  GCGGCTAAAT GCTGCAGGAT AG

```

The PSORT algorithm predicts an outer membrane location (0.940).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 66A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 66B) and for FACS analysis. A his-tagged protein was also expressed.

These experiments show that cp6890 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 67

The following *C.pneumoniae* protein (PID 6172323) was expressed <SEQ ID 133; cp0018>:



```

351  GGAAGATGCG TCCTCCGTAT ATCTCTATGC GTTTTTACCT ATCAAAAATG
401  CTCGGGCAAT CTTTGATGAT ACTGAGTCTC CAGAAAATCT AGGAGTCCGA
451  GCTTTAGATA AGCGTCATCT CGAAATTCAG TTAGAACTC CCTGCGCGCA
501  TTTCTACAT TTCTTGACTC TTCCTATTTT TTTCCCTGTT CATGAAACTC
551  TGCGAAACTA TAGCACCTCT TTTGAAGAGA TGCCCATFAC CTGCGGTGCT
601  TTCCGCCCTG TGTCTCTAGA AAAAGGCCTG AGACTCCATC TAGAGAAAAA
651  CCCTATGTAC CATAATAAAA GCCGTGTGAA ACTACATAAA ATTATTGTAC
701  AGTTTATCTC AAACGCTAAC ACTGCAGCCA TTCTATTCAA ACATAAGAAA
751  TTAGATTGGC AAGGACCTCC TTGGGGAGAA CCTATCCCTC CAGAAATCTC
801  AGCTTCTCTA CATCAAGATG ACCAGCTCTT TTCTCTTCCG GCGGCTTCGA
851  CTACATGGTT ACTCTTTAAT ATACAAAAAA AACCTTGGA CAATGCTAAA
901  TTACGCAAGG CATTGAGCCT TGCAATAGAC AAAGATATGT TAACCAAAGT
951  GGTATACCAA GGTCTTGCA GAACTACAGA TCATATCCTA CATCCAAGAC
1001 TTTATCCAGG GACCTATCCC GAACGGAAAA GACAAAACGA AAGAATTCTT
1051 GAGGCTCAAC AACTCTTTGA AGAAGCTCTA GACGAACCTC AAATGACACG
1101 CGAAGATCTA GAAAAGGAAA CTTTGACTTT CTCAACCTTT TCTTTTTCTT
1151 ACGGAAGGAT TTGCCAAATG CTAAGAGAAC AATGGAAGAA AGTCTTAAAA
1201 TTTACTATCC CTATAGTAGG CCAAGAGTTT TTCACAATAC AAAAAAACTT
1251 CCTAGAGGGG AACTATTCCC TAACCGTGAA CCAATGGACC GCAGCATTTA
1301 TTGATCCGAT GTCTTATCTC ATGATCTTTG CCAATCCTGG AGGAATTTCC
1351 CCCTATCACC TCCAAGATTC AACTTTTCAA ACTCTTCTCA TAAAGATCAC
1401 TCAAGAACAT AAAAAACACC TACGAAATCA GCTTATTATT GAAGCCCTTG
1451 ACTATTTAGA AACTGTTCAC ATTCTCGAAC CACTATGTCA TCCAAATCTT
1501 CGAATTGCTT TGAACAAAAA CATTAAAAAC TTTAATCTTT TTGTTGACG
1551 AACTTCAGAC TTTCGTTTTA TAGAAAAACT ATAG

```

The PSORT algorithm predicts an outer membrane lipoprotein (0.790).

The protein was expressed in *E.coli* and purified as a his-tag product and a GST-fusion protein, as shown in Figure 64A. The recombinant his-tag protein was used to immunise mice, whose sera were used in a Western blot (Figure 64B). The recombinant GST-fusion protein was also used to immunise mice, whose sera were used in a Western blot (Figure 64C) and for FACS analysis (Figure 64D).

These experiments show that cp6467 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 65

The following *C.pneumoniae* protein (PID 4376679) was expressed <SEQ ID 129; cp6679>:

```

1  MRKMLVLLAS LGLLSPTLSS CTHLGSSGSY HPKLYTSGSK TKGVIAMLPV
51  FHRPGKSLEP LPWNLQGEFT EBISKRFYAS EKVFLIKHNA SPQTVSQFYA
101 PIANRLPETI IEQFLPAEFI VATELLEQKT GKEAGVDSVT ASVRVRVFDI
151 RHHKIALIYQ EIIECSQPLT TLVNDYHRYG WNSKHFDSTP MGLMHSRLFR
201 EVVARVEGYV CANYS*

```

A predicted signal peptide is highlighted.

The cp6679 nucleotide sequence <SEQ ID 130> is:

```

1  ATGCGAAAAA TGTTGGTATT ATTGGCATCT TTAGGACTTC TATCCCCAAC
51  CCTATCCAGC TGCACTCACT TAGGCTCTTC AGGAAGTTAT CATCCTAAGC
101 TATACACTTC AGGGAGCAAA ACTAAAGGTG TGATTGCGAT GCTTCCTGTA
151 TTTTCATCGCC CAGGAAAGAG TCTTGAACCT TTACCTTGGG ACCTCCAAGG
201 AGAATTTACT GAAGAGATCA GCAAAAAGGT TTAGGCTTCG GAAAAGGTCT
251 TCCTGATCAA GCACAATGCT TCACCTCAGA CAGTCTCTCA GTTCTATGCT
301 CCGATTGCGA ATCGTCTACC CGAAACAATT ATTGAGCAAT TTCTTCCTGC
351 AGAATTCATT GTTGCTACAG AACTGTTAGA ACAAAGACA GGGAAAGAAG
401 CAGGTGTCGA TTCTGTAACA GCGTCTGTAC GTGTTGCGGT TTTTGATATC
451 CGTCATCATA AAATAGCTCT CATTTATCAA GAGATTATCG AATGCAGCCA
501 GCCTTTAACT ACCCTAGTCA ATGATTATCA TCGCTATGGC TGGAACTCAA
551 AACATTTTGA TTCAACGCC ATGGGCTTAA TGCATAGCCG TCTTTTCCGG

```

-108-

301 DDIDEESIRL QQEAEAAALR LPEEMSAFEG YIKVVEHLE NMKSLPYDGH  
 351 GLEEKTKHQI RVVRSSLKAM VPEFLDIRRI FEEFFFLS ARKRLIDLAT  
 401 TLVERKILTE QLERNNLKKA FSYLQDSIF KKIIDNFEKL AWKFMILSKS  
 451 ICRFTIIFEN HEHGVAKSLL HKNVLLLEKV IYRSLOKSYR DIGMSSAKMK  
 501 ILHGNPFFSL EDNKKTIMKE HAEMLESLS YRKVFLALSD ENVVDTSPDP  
 551 KKWDLGIPC RDALSEISRD EQWQKKAHLK HQESLYTQAR DRLTDQSSKE  
 601 NQKELEKAEQ EYISSWERVK KFEIERVQER IRAIQKLYPN ILEREETTG  
 651 QETVTPVQG TTASSDLTDI LGRIEVSSRE DNQNEQSCVK VLSHEVEMS  
 701 WEVKQEYGP KKEFQDQMG LERFFTEHIE ELEVLQKDYS KHLSYFKKVN  
 751 NKKEVQYAKF RLKLVESDLE GILAQTESAE SLLTQEELPI LATRGALEKA  
 801 VFKGSLCCAL ASKAKPYFEE DPRFQSDTQ LRALTRLRQE AKASLEEBEIK  
 851 RFSNLENDIA EERRLLKESK QTFERAGLV LREIAVESTY DLRSLTNTWE  
 901 GTPSEKQVYF SMYLNYYNEE KRRAKTRLVE MTQRYRDFKM ALEAMQFNEE  
 951 ALLQEELSIQ APSE\*

A predicted signal peptide is highlighted.

The cp6262 nucleotide sequence <SEQ ID 136> is:

1 ATGAGGAAAC TTCGTATTCT TCGCATCGTT CTCATAGCTT TGAGCATTAT  
 51 TTTGATTGCA GGTGGTGTGG TATTGCTTAC TGAGCGATC CCTGGATTAA  
 101 GTTCAGTCAT TTCTTCCCCG GCAGGGATGG GTGCCTGTGC TTTGGGATGT  
 151 GTGATGCTTG CTTTAGGGAT CGATGTTCTT CTGAAGAAAC GAGAAGTCCC  
 201 TATAGTTCTC GCATCTGTAA CTACGACACC AGGAACGGC AGCCCTAGAA  
 251 GTGGTATTTT TATTTTCAGGA GCTGATAGCA CCATACGTTT TCTTCTTACG  
 301 TATCTCTTGG ACGAGGGACA TCCACAATCC ATGAGGAAAC TTCGTATTCT  
 351 TCGCATCGTT CTCATAGTTT TTAGCATTAT TTTGATTGCA AGTGGTGTGG  
 401 TATTGCTTAC TGAGCGATC CCTGGATTAA GTTCAGTCAT TTCTTCCCCG  
 451 GCAGGGATGG GTGCCTGTGC TTTGGGATGT GTGATGCTTG CTTTAGGGAT  
 501 CGATGTTCTT CTGAAGAAAC GAGAAGTCCC TATAGTTCTC GCATCTGTAA  
 551 CTACGACACC AGGAACGGC AGCCCTAGAA GTGGTATTTT TATTTTCAGGA  
 601 GCTGATAGCA CCATACGTTT TCTTCTTACG TATCCCTTGG ACGAGGGACA  
 651 TCCACAATCC ATGAGGAAAC TTCGTATTCT TCGCATCGTT CTCATAGTTT  
 701 TTAGCATTAT TTTGATTGCA AGTGGTGTGG TATTGCTTAC TGAGCGATC  
 751 CCTGGATTAA GCTCGATCAT TTCTTCCCCA GCGGAGATGG GTGCTTGTGC  
 801 TTTGGGATGT GTGATGCTTG CTTTGGGGAT CGACGTTCTT CTGAAGAAAC  
 851 GAGAAGTCCC TATAGTAGTT CCCGCACCTA TTCCTGAAGA AGTCGTCATA  
 901 GATGATATAG ATGAAGAGAG TATACGGCTG CAGCAGGAAG CTGAAGCCCG  
 951 TTTGCAAGA CTTCTGAGG AGATGAGTGC ATTTGAAGGT TACATAAAAG  
 1001 TTGTCGAGAG TCATTTGGAG AACATGAAAA GCCTGCCTTA TGATGGTCAAT  
 1051 GGGCTAGAAG AGAAAACGAA ACATCAGATA AGAGTCGTCA GATCTTCTTT  
 1101 GAAGGCTATG GTTCCAGAAT TTTTAGATAT CAGAAGAATT TTTGAAGAAG  
 1151 AAGAGTTCTT TTTTCTCTCA GCTCGCAAAC GACTTATAGA TTTAGCTACT  
 1201 ACTTTAGTAG AGAGAAAAAT TTTAACAGAG CAACTTGAGC GCAATAATTT  
 1251 AAGGAAAGCG TTTTCTTATT TATATCAGGA CTCAATTTTT AAAAAATTA  
 1301 TTGATAACTT CGAGAAGTTA GCATGGAAAT TTATGATTTT GAGTAAATCA  
 1351 ATTTGTGCGT TTACAATTAT TTTGAAAAT CATGAACATG GTGTAGCAAA  
 1401 GAGCCTGTTA CACAAGAATG CAGTGTACT GGAGAAGGTA ATCTATAGGA  
 1451 GTTTGCAAAA AAGCTATAGA GATATAGGCA TGTCATCTGC AAAGATGAAA  
 1501 ATCTTGACAG GCAACCCTTT TTCTCTTTG GAAGATAATA AAAAGACGAT  
 1551 AATGAAAGAA CACGCAGAGA TGCTGAAAG TCTCAGTAGC TATAGGAAGG  
 1601 TATTTTGTAG TCTATCTGAT GAGAACGTTG TAGATACACC TAGCGATCCA  
 1651 AAGAAATGGG ATTTGTGAGG AATCCCCTGT AGGGACGCGT TGTCTGAGAT  
 1701 TTCTCGTGAT GAACAGTGGC AGAAGAAAGC ACATCTAAAG CATCAAGAGT  
 1751 CCCTCTATAC GCAAGCTAGG GATCGTTTAA CAGACCAGAG CTCTAAAGAA  
 1801 AATCAGAAAG AGTTAGAGAA AGCTGAACAA GAGTACATAT CTCTTTGGGA  
 1851 ACGGGTTAAA AAATTTGAGA TTGAGAGAGT ACAGGAGAGG ATACGGGCAA  
 1901 TTCAAAGCTT TTATCCTAAT ATCCTCGAGA GAGAAGAAGA AACCACAGGT  
 1951 CAGGAGACTG TGACTCCAAC TGTTCAGGG ACGACGGCTT CATCCGATTT  
 2001 AACAGATATT TTAGGAAGAA TAGAGGTCTC CAGTAGGGAG GATAATCAGA  
 2051 ATCAAGAGTC TTGTGTAAAA GTCTTAAGAA GTCATGAGGT AGAAATGAGC  
 2101 TGGGAAGTCA AACAAGAGTA TGGCCCTAAG AAAAAAGAAT TTCAGGATCA  
 2151 AATGGGTTCT TTAGAGAGGT TTTTACAGA GCATATTGAA GAGTTAGAAG  
 2201 TATTACAGAA GGACTACTCT AAACACTTGT CTTATTTTAA AAAAGTAAAC  
 2251 AATAAGAAAG AGGTTCAATA TGCGAAGTTT AGGTTGAAGG TTTTAGAGTC  
 2301 AGATTTAGAA GGGATTCTAG CTCAGACTGA GAGTGCTGAG AGTCTGTATA  
 2351 CTCGAAGAAG ACTTCCGATT CTGCAACTC GGGGAGCCTT AGAGAAAGCT  
 2401 GTTTTCAAAG GGAGTCTATG TTGCGCGCTA GCAAGCAAAG CAAAACCCTA

```

1  MKTSVSMLLA  LLCSGASSIV  LHAATTPPLNP  EDGFIFEGNT  NTFSPKSTTD
51 AAGTTYSLTG  EVLYIDPGKG  GSITGTCFVE  TAGDLTFLGN  GNTLKFLSVD
101 AGANIAVAHV  QGSKNLSFTD  FLSLVITESP  KSAVTTGKGS  LVSLGAVQLQ
151 DINTLVLTSTN  ASVEDGGVIK  GNSCLIQGIK  NSAIFGQNTS  SKKGGAIIST
5  201 QGLTIENNLG  TLKFENENKAV  TSGGALDLGA  ASTFTANHEL  IFSQNKTSN
251 AANGGAINCS  GDLTFTDNTS  LLLQENSTMQ  DGGALCSTGT  ISITGSDSIN
301 VIGNTSGQKG  GAISAASLKI  LGGQGGALFS  NNVVTHATPL  GGAIFINTGG
351 SLQLFTQGGD  IVFEGNQVTT  TAPNATTKRN  VIHLESTAKW  TGLAASQGNA
10 401 IYFYDPITTN  DTGASDNLRI  NEVSANQKLS  GSIVFSGERL  STAEAIAENL
451 TSRINQPVTL  VEGSLVLKQG  VTLITQGFSS  EPESTLLLDL  GTSL*

```

A predicted signal peptide is highlighted.

The cp0018 nucleotide sequence <SEQ ID 134> is:

```

1  ATGAAGACTT  CAGTTTCTAT  GTTGTTGGCC  CTGCTTTGCT  CGGGGGCTAG
51 CTCTATTGTA  CTCCATGCCG  CAACCACTCC  ACTAAATCCT  GAAGATGGGT
15 101 TTATTGGGGA  GGGCAATACA  AATACTTTT  CTCCGAAATC  TACAACGGAT
151 GCTGCAGGAA  CTACCTACTC  TCTCACAGGA  GAGGTTCTGT  ATATAGATCC
201 GGGGAAAGGT  GGTTC AATTA  CAGGAACCTG  CTTTGTAGAA  ACTGCTGGCG
251 ATCTTACATT  TTTAGGTAAT  GGAAATACCC  TAAAGTTCCT  GTCGGTAGAT
20 301 GCAGGTGCTA  ATATCGCGGT  TGCTCATGTA  CAAGGAAGTA  AGAATTTAAG
351 CTTACACAGAT  TTCCTTTCTC  TGGTGATCAC  AGAATCTCCA  AAATCCGCTG
401 TTACTACAGG  AAAAGGTAGC  CTAGTCAGTT  TAGGTGCAGT  CCAACTGCAA
451 GATATAAACA  CTCTAGTTCT  TACAAGCAAT  GCCTCTGTCG  AAGATGGTGG
501 CGTGATTAAA  GGAACTCCT  GCTTGATTCA  GGAATCAAA  AATAGTCCGA
25 551 TTTTTGGACA  AAATACATCT  TCGAAAAAAG  GAGGGGCGAT  CTCCACGACT
601 CAAGGACTTA  CCATAGAGAA  TAACTTAGGG  ACGCTAAAGT  TCAATGAAAA
651 CAAAGCAGTG  ACCTCAGGAG  GCGCCTTAGA  TTTAGGAGCC  GCGTCTACAT
701 TCACTGCGAA  CCATGAGTTG  ATATTTTCAC  AAAATAAGAC  TTCTGGGAAT
751 GCTGCAATG  GCGGAGCCAT  AAATTGCTCA  GGGGACCTTA  CATTTACTGA
801 TAACACTTCT  TTGTTACTTC  AAGAAAATAG  CACAATGCAG  GATGGTGGAG
30 851 CTTTGTGTAG  CACAGGAACC  ATAAGCATTA  CCGGTAGTGA  TTCTATCAAT
901 GTGATAGGAA  ATACTTCAGG  ACAAAAAGGA  GGAGCGATTT  CTGCAGCTTC
951 TCTCAAGATT  TTGGGAGGGC  AGGGAGGCGC  TCTCTTTTCT  AATAACGTAG
1001 TGACTCATGC  CACCCCTCTA  GGAGGTGCCA  TTTTATCAA  CACAGGAGGA
35 1051 TCCTTGCAAG  TCTTCACTCA  AGGAGGGGAT  ATCGTATTCG  AGGGGAATCA
1101 GGTCACTACA  ACAGCTCCAA  ATGCTACCAC  TAAGAGAAAT  GTAATTCACC
1151 TCGAGAGCAC  CGGAAGTGG  ACGGGAAGTG  CTGCAAGTCA  AGGTAACGCT
1201 ATCTATTTCT  ATGATCCCAT  TACCACCAAC  GATACGGGAG  CAAGCGATAA
1251 CTTACGTATC  AATGAGGTCA  GTGCAAATCA  AAAGCTCTCG  GGATCTATAG
1301 TATTTTCTGG  AGAGAGATTG  TCGACAGCAG  AAGCTATAGC  TGAAAATCTT
40 1351 ACTTCGAGGA  TCAACCAGCC  TGTCACCTTA  GTAGAGGGGA  GCTTAGTACT
1401 TAAACAGGGA  GTGACCTTGA  TCACACAAGG  ATTCTCGCAG  GAGCCAGAA
1451 CCACGCTTCT  TTTGGATCTG  GGGACCTCAT  TATAA

```

The PSORT algorithm predicts outer membrane (0.935).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 67A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 67B) and for FACS analysis.

These experiments show that cp0018 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 68

The following *C.pneumoniae* protein (PID 4376262) was expressed <SEQ ID 135; cp6262>:

```

1  MRKLRLAIV  LIALSIILIA  GGVLLTVAI  PGLSSVISSP  AGMGACALGC
51 VMLALGIDVL  LKKREVPIVL  ASVTTTPGTG  SPRSGISISG  ADSTIRSLPT
101 YLLDEGHPQS  MRKLRLAIV  LIVFSIILIA  SGVLLTVAI  PGLSSVISSP
151 AGMGACALGC  VMLALGIDVL  LKKREVPIVL  ASVTTTPGTG  SPRSGISISG
55 201 ADSTIRSLPT  YPLDEGHPQS  MRKLRLAIV  LIVFSIILIA  SGVLLTVAI
251 PGLSSIISP  AEMGACALGC  VMLALGIDVL  LKKREVPIV  PAPIPEEVVI

```

-110-

```

1401 AAAGCTTTGC TCTCTACGTC TTGATGAAAA AGAGTTATTA CAAAAAGAAA
1451 TCAAGAAAGA GGAATTTTAT CAAAAGAAAC AACAAAGGCA TGCAGATAGA
1501 TCACGTCATA CTACGTATCA AAAGCTACGA ATTGCTGAAG AGCTTGCTCT
1551 TGAGCTGAAG AAGAAAATCT AA

```

5 The PSORT algorithm predicts cytoplasmic location (0.412).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 69A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 69B) and for FACS analysis.

10 These experiments show that cp6269 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 70

The following *C.pneumoniae* protein (PID 4376270) was expressed <SEQ ID 139; cp6270>:

```

15 1  MKIPLRFLLI  SLVPTLSMSN  LLGAATTEEL  SASNSFDGTT  STTSFSSKTS
    51  SATDGTNYVF  KDSVVIENVP  KTGETQSTSC  FKNDAAAGDL  NFLGGGFSFT
    101  FSNIDATTAS  GAAIGSEAA  KVTTLSGFSA  LSFLKSPAST  VTNGLGAINV
    151  KGNLSLLDND  KVLIQDNFST  GDGGAINCAG  SLKIANNKSL  SFIGNSSSTR
    201  GGAIHTKNLT  LSSGGETLFQ  GNTAPTAAGK  GGAIAIADSG  TLSISGDSGD
    251  IIFEGNTIGA  TGTVSHSAID  LGTSAKITAL  RAAQGHTIYF  YDPITVTGST
    301  SVADALNINS  PDTGDNKEYT  GTIVFSGEKL  TEAEAKDEKN  RTSKLLQNVA
    351  FKNGTVVLKG  DVVLSANGFS  QDANSKLMD  LGTSLVANTE  SIELTNLEIN
    401  IDSLRNGKKI  KLSAATAQKD  IRIDRPVULA  ISDESFYQNG  FLNEDHSDYD
    451  ILELDAGKDI  VISADSRSID  AVQSPYGYQG  KWTINWSTDD  KKATVSWAQ
    501  SFNPTAEQEA  PLVPNLLWGS  FIDVRSFQNF  IELGTEGAPY  EKRFWVAGIS
    551  NVLHRSGREN  QRKFRHVSOG  AVVGASTRMP  GGDTLSLGFA  QLFARDKDYF
    601  MNTNFAKTYA  GSLRLQHDAS  LYSVVSILLG  EGGLREILLP  YVSKTLPCSF
    651  YGQLSYGHTD  HRMKTESLPP  PPPTLSTDHT  SWGGYVWAGE  LGTRVAVENT
    701  SGRGFFQEYT  PFVKVQAVYA  RQDSFVELGA  ISRDFSDSL  YNLAIPLGIK
    751  LEKRFAEQYY  HVVAMYSPOV  CRSNPKCTTT  LLSNQGSWKT  KGSNLARQAG
    801  IVQASGFRSL  GAAAEFLGNF  GFEWRGSSRS  YNVDAGSKIK  F*

```

30 A predicted signal peptide is highlighted.

The cp6270 nucleotide sequence <SEQ ID 140> is:

```

35 1  ATGAAGATTC  CACTCCGCTT  TTTATTGATA  TCATTAGTAC  CTACGCTTTC
    51  TATGTGCAAT  TTATTAGGAG  CTGCTACTAC  CGAAGAGTTA  TCGGCTAGCA
    101  ATAGCTTCGA  TGGAACACAA  TCAACAACAA  GCTTTTCTAG  TAAAACATCA
    151  TCGGCTACAG  ATGGCACCAA  TTATGTTTTT  AAAGATTCTG  TAGTTATAGA
    201  AAATGTACCC  AAAACAGGGG  AAACCTCAGT  TACTAGTTGT  TTTAAAAATG
    251  ACGCTGCAGC  TGGAGATCTA  AATTTCTTAG  GAGGGGGATT  TTCTTTCACA
    301  TTTAGCAATA  TCGATGCAAC  CACGGCTTCT  GGAGCTGCTA  TTGGAAGTGA
    351  AGCAGCTAAT  AAGACAGTCA  CGTTATCAGG  ATTTTCGGCA  CTTTCTTTTC
    401  TTAAATCCCC  AGCAAGTACA  GTGACTAATG  GATTGGGAGC  TATCAATGTT
    451  AAAGGGAATT  TAAGCCTATT  GGATAATGAT  AAGGTATTGA  TTCAGGACAA
    501  TTTCTCAACA  GGAGATGGCG  GAGCAATTAA  TTGTGCAGGC  TCCTTGAAGA
    551  TCGCAAACAA  TAAATCCCTT  TCTTTTATG  GAAATAGTTC  TTCAACACGT
    601  GCGGAGCGA  TTCATACCAA  AAACCTCACA  CTATCTTCTG  GTGGGGAAC
    651  TCTATTTTCA  GGAATACAG  CGCCTACGGC  TGCTGGTAAA  GGAGGTGCTA
    701  TCGCGATTGC  AGACTCTGGC  ACCCTATCCA  TTTCTGGAGA  CAGTGGCGAC
    751  ATTATCTTTG  AAGGCAATAC  GATAGGAGCT  ACAGGAACCG  TCTCTCATAG
    801  TGCTATTGAT  TTAGGAACCT  GCGCTAAGAT  AACTGCGTTA  CGTGCTGCGC
    851  AAGGACATAC  GATATACTTT  TATGATCCGA  TTAAGTGAAC  AGGATCGACA
    901  TCTGTTGCTG  ATGCTCTCAA  TATTAATAGC  CCTGATACTG  GAGATAACAA
    951  AGAGTATACG  GGAACCATAG  TCTTTTCTGG  AGAGAAGCTC  ACGGAGGCG
    1001  AAGCTAAAGA  TGAGAAGAAC  CGCACTTCTA  AATTACTTCA  AAATGTTGCT
    1051  TTTAAAAATG  GGACTGTAGT  TTTAAAGAGT  GATGTCGTTT  TAAGTGGGAA
    1101  CGGTTTCTCT  CAGGATGCAA  ACTCTAAGTT  GATTATGGAT  TTAGGGACGT
    1151  CGTTGGTTGC  AAACACCGAA  AGTATCGAGT  TAACGAATTT  GGAAATTAAT
    1201  ATAGACTCTC  TCAGGAACGG  GAAAAAGATA  AAACCTCAGT  CTGCCACAGC

```

```

2451 TTTTGAAGAG GATCCCAGAT TCCAAGATTC TGATACGCAA TTGCGAGCTC
2501 TGA CTCTAAG GTTACAGGAG GCTAAGGCAA GCCTGGAAGA AGAGATAAAG
2551 AGATTTTCAA ATCTTGAGAA CGATATTGCA GAGGAAAGAC GCCTTCTTAA
2601 AGAGAGCAAAG CAGACGTTTCG AAAGAGCAGG TTTAGGGGTT CTCCGAGAAA
2651 TTGCAGTCGA GTCTACTTAT GATTTGCGTT CCTTAACAAA TACATGGGAA
2701 GGGACCCAG AGAGTGAGAA GGTCTATTTT AGCATGTATC TTAATTATTA
2751 CAACGAAGAG AAACGTAGGG CTAAAACAAG ATTGGTTGAA ATGACACAGA
2801 GGTATAGAGA TTTTAAATG GCCTTGGGAA CTATGCAGTT TAATGAAGAA
2851 GCCCTTTTGC AAGAGGAACT CTCTATTCAA GCTCCAGTG AATAA

```

10 The PSORT algorithm predicts inner membrane (0.660).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 68A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 68B) and for FACS analysis.

15 These experiments show that cp6262 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 69

The following *C.pneumoniae* protein (PID 4376269) was expressed <SEQ ID 137; cp6269>:

```

1 MYQENLRLLLE RLLYSNVQKS YADRLFSYEK TKMVHDTPLI PWEEDEKKECA
51 EA EKAFLEQQ KILLDYGKSI FWLNENDEIN LNDPWSWGLN TVRTRKVFQE
101 VDDSERWNHK VLIQKLEDDY EKLLEESSKE STEANKLLS DLVDRLEDAK
151 TKFFFLKQEE VETRVKDLRA RYGGTVDPKQ DTEAKKKVEL EASLETFLDS
201 IESELVQCLE DQDIYWKEQD VKDLARTQEL EEQDIEAKRE EAAEDLRSLN
251 ERLKKSKTML DRAKWHIENA EDSITWWT SQ IEMKDMKARL KILKEDITSV
301 LPEIDEIETC LSLEELPLLT TRELLTKSYL KFKICSETLL KMTSVFENNI
25 351 YVQYEYVQLQ NLGFKLQGIS QRFGKKQDDF ANLEEQVALQ KKRLRELTON
401 FEIQGFNFMK EDFKAAAKDL YIRSTAEQKM NFDVPCMELE RRYHEEVNKP
451 LLELMYNCAD SYRDAKKKLC SLRLDEKELL QKEIKKEEFY QKKQQRHADR
501 SRHTTYQKLR IAEELALELK KKI*

```

The cp6269 nucleotide sequence <SEQ ID 138> is:

```

30 1 ATGTACCAGG AGAATCTAAG ATTGTTGGAA AGGCTTCTTT ATAATAGTGT
51 TCAAAAGAGC TATGCGGATC GGCTGTTTTT CTATGAAAAG ACAAGATGAG
101 TGCACGATAC TCCGCTGATT CCTTGGGAAG AGGATAAGGA AAAATGTGCT
151 GAAGCTGAGA AAGCTTTCTT AGAGCAACAG AAGATTCTCC TAGATTATGG
201 AAAATCTATC TTTTGGCTGA ATGAGAACGA TGAGATCAAT TTAAACGATC
35 251 CTTGGAGTTG GGGTCTTAAT ACGGTGAGGA CTAGGAAAGT ATTCCAAGAG
301 GTTGACGACA GTGAACGTTG GAATCATAAG GACTCATTC AAAAAGCTCGA
351 GGACGATTAT GAGAAAGTTC TAGAGGAAAG TTCAAAAGAG TCTACTGAAG
401 CAAATAAGAA GCTTTTATCT GACTTAGTAG ATCGTCTTGA AGATGCTAAG
451 ACAAATTTT TCCTGAAGAA ACAGGAGGAG GTGGAGACTC GCGTTAAGGA
40 501 TCTTAGAGCT CGATATGGAG GCACAGTAGA TCCTAAGCAG GATACGGAAG
551 CTAAGAAGAA AGTCAATTG GAGGCTAGCT TAGAAACCTT TTTAGATTCC
601 ATCGAATCAG AGCTAGTACA GTGTTTAGAA GATCAAGATA TATATTGGAA
651 AGAACAGGAT GTCAAAGATC TAGCACGTAC GCAAGAGCTC GAGGAACAAG
701 ATATTGAAGC GAGAGGGGAA GAAGCTGCCG AAGACCTAAG AAGTCTTAAT
45 751 GAGCGTTTAA AGAAGTCAAA AACTATGTTA GATAGGGCTA AATGGCATAT
801 TGAAATGCT GAGGACAGTA TTACCTGGTG GACTAGTCAG ATAGAAATGA
851 AGGATATGAA AGCAAGACTG AAGATCTTAA AAGAAGATAT AACAAAGTGT
901 CTACCTGAAA TAGATGAGAT TGAAACGTGT TTAAGCTTAG AGGAGCTTCC
951 TTTGCTTACG ACCAGGGAAC TCTTAACATA GTCCTACCTA AAGTTTAAGA
50 1001 TTTGTTCCGA AACACTATTA AAAATGACTT CTGTGTTTGA GAACAATATC
1051 TAGTTTCAGG AGTACGAGGT TCAGCTGCAA AATCTAGGCT TTAAGTTACA
1101 AGGTATATCT CAGAGATTCG GAAAGAAACA AGACGATTTT GCGAATCTAG
1151 AGGAACAGGT TGCTTTGCAA AAGAAACGAC TCAGAGAGCT CACTCAGAAT
1201 TTTGAAATAC AAGGATTCAA TTTCATGAAA GAAGATTTTA AGGCAGCCGC
55 1251 TAAAGATCTT TATATAAGAA GTACAGCTGA ACAAAGATG AACTTTGATG
1301 TGCCTTGCAAT GGAGCTCTTC CGTAGGTATC ATGAGGAGGT CAACAAGCCG
1351 CTTCTTGAGT TGATGTACAA TTGTGCAGAC AGTTATAGAG ATGCTAAGAA

```

701 TAAAAAGCGA ATTTCTTATT TCCACAACCT TTATAGATAC GGCCAACCCC  
751 TTCTAA

The PSORT algorithm predicts cytoplasmic (0.158).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 71A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 71B) and for FACS analysis.

These experiments show that cp6402 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 72

10 The following *C.pneumoniae* protein (PID 4376520) was expressed <SEQ ID 143; cp6520>:

1 MKHYLSFSPS ADFFSKQGA I ETQVLFGERV LVKGSTCYAY SQLFHNELLW  
51 KPYPGHSFRS TLVPCTPEFH IHPNVSVVSV DAFLDPWGIP LPFGTLLHVN  
101 SQNTVIFPKD ILNHMNTIWG SGTPQC DPRH LRRLNYNFFA ELLIKDADLL  
151 LNFYPVWGGR SVHESLEKPG VDCSGFINIL YQAQGYNVPR NAADQYADCH  
15 201 WISSFENLPS GGLIFLYPKE EKRIHVMLK QDSSTLIHAS GGGKKVEYFI  
251 LEQDGKFLDS TYLFFRNNQR GRAFFGIPRK RKAFL\*

The cp6520 nucleotide sequence <SEQ ID 144> is:

1 ATGAAACACT ACCTATCATT TTCTCCTTCT GCTGATTTTT TCTCTAAACA  
51 GGGTGCATATT GAAACTCAAG TCCTTTTGG AGAGCGCGTC TTAGTCAAAG  
20 101 GGAGCACCTG CTATGCATAT TCCCAATTAT TCCACAATGA GCTGTATATGG  
151 AAGCCCTATC CAGGTCATAG CTTTCGTCT ACCCTAGTCC CCTGCACTCC  
201 TGAATTTTCAT ATCCATCCAA ATGTTTCTGT GGTTCCTGTG GATGCATTTT  
251 TAGATCCTTG GGGGATCCCT CTTCTTTTG GAACTTTACT CCATGTGAAT  
301 TCTCAAAATA CCGTTATTTT CCCTAAGGAT ATTCTCAATC ATATGAACAC  
25 351 CATCTGGGCG TCCGGCACAC CTCAATGCGA TCCTAGACAT CTACGTCGTC  
401 TAAATTATAA CTTCTTTGCT GAACTTTTAA TTAAAGACGC AGACCTTTTA  
451 CTGAACCTTC CCTATGTATG GGGAGGACGG TCTGTACACG AAAGTCTGGA  
501 AAAGCCGGGT GTTGATTGTT CGGGATTAT CAATATCCTT TACCAGGCAC  
551 AGGGATACAA CGTCCCTAGA AACGCTGCAG ATCAATATGC GGATTGTCAT  
30 601 TGGATCTCTA GCTTTGAGAA CCTTCCTTCT GGTGGGTAA TATTTCTTTA  
651 CCTAAAGAA GAAAAGCGTA TTTCTCATGT TATGTTGAAA CAGGATAGTT  
701 CCACCCTCAT TCATGCTTCT GGTGGAGGGA AAAAAGTGA GTATTTTCATT  
751 TTAGAACAAG ATGGGAAGTT TTTAGATTCG ACTTATCTAT TTTT TAGAAA  
801 TAATCAGAGG GGACGGCAT TTTTGGGAT CCCTAGAAA AGAAAAGCCT  
35 851 TTCTGTAA

The PSORT algorithm predicts cytoplasmic (0.265).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 72A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 72B) and for FACS analysis.

40 These experiments show that cp6520 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 73

The following *C.pneumoniae* protein (PID 4376567) was expressed <SEQ ID 145; cp6567>:

1 MTSPPIFQSS GDASFLAEQP QQLPSTSESQ LVTQLLTMMK HTQALSETVL  
45 51 QQQRDLRPTA SIILQVGGAP TGGAGAPFQP GPADDHHPPI PPPVVPAPQE  
101 TEITIRSEL QLMRSTLQQS TKGARTGVLV VTAILMTISL LAIIIIILAV  
151 LGFTGVLPOV ALLMQGETNL IWAMVSGSII CFIALIGTLG LILTNNKNTPL

-111-

1251 TCAGAAAGAT ATTCGTATAG ATCGTCCTGT TGTACTGGCA ATTAGCGATG  
 1301 AGAGTTTTTA TCAAAATGGC TTTTGAATG AGGACCATTC CTATGATGGG  
 1351 ATTCTTGAGT TAGATGCTGG GAAAGACATC GTGATTTCTG CAGATTCCTG  
 1401 CAGTATAGAT GCTGTACAAT CTCCGTATGG CTATCAGGGA AAGTGGACGA  
 1451 TCAATTGGTC TACTGATGAT AAGAAAGCTA CCGTTTCTTG GCGGAAGCAG  
 1501 AGTTTTAATC CCACTGCTGA GCAGGAGGCT CCGTTAGTTC CTAATCTTCT  
 1551 TTGGGGTTCT TTTATAGATG TTCGTTCCCTT CCAGAATTTT ATAGAGCTAG  
 1601 GTACTGAAGG TGCTCCTTAC GAAAAGAGAT TTTGGGTTGC AGGCATTTCC  
 1651 AATGTTTTTG ATAGGAGCGG TCGTGAAAAT CAAAGGAAAT TCCGTCATGT  
 1701 GAGTGGAGGT GCTGTAGTAG GTGCTAGCAC GAGGATGCCG GGTGGTGATA  
 1751 CCTTGTCTCT GGGTTTTGCT CAGCTCTTTG CGCGTGACAA AGACTACTTT  
 1801 ATGAATACCA ATTTGCGAAA GACCTACGCA GGATCTTTAC GTTTGCAGCA  
 1851 CGATGCTTCC CTATACTCTG TGGTGAGTAT CCTTTTAGGA GAGGGAGGAC  
 1901 TCCGCGAGAT CCTGTTGCCT TATGTTTCCA AGACTCTGCC GTGCTCTTTC  
 1951 TATGGGCAGC TTAGCTACGG CCATACGGAT CATCGCATGA AGACCGAGTC  
 2001 TCTACCCCCC CCCCCCCCGA CGCTCTCGAC GGATCATACT TCTTGGGGAG  
 2051 GATATGTCTG GGCTGGAGAG CTGGGAACTC GAGTTGCTGT TGAAAATACC  
 2101 AGCGGCAGAG GATTTTTCCA AGAGTACACT CCATTTGTAA AAGTCCAAGC  
 2151 TGTTTACGCT CGCCAAGATA GCTTTGTAGA ACTAGGAGCT ATCAGTCGTG  
 2201 ATTTTAGTGA TTCGCATCTT TATAACCTTG CGATTCCTCT TGGAAACAAG  
 2251 TTAGAGAAAC GGTTCGAGA GCAATATTAT CATGTTGTAG CGATGTATTC  
 2301 TCCAGATGTT TGTCGTAGTA ACCCCAAATG TACGACTACC CTACTTTCCA  
 2351 ACCAAGGGAG TTGGAAGACC AAAGGTTCTGA ACTTAGCAAG ACAGGCTGGT  
 2401 ATTGTTTCAGG CCTCAGGTTT TCGATCTTTG GGAGCTGCAG CAGAGCTTTT  
 2451 CGGGAAC'TTT GGCTTTGAAT GCGGGGATC TTCTCGTAGC TATAATGTAG  
 2501 ATGCGGGTAG CAAAATCAA TTTTAG

The PSORT algorithm predicts outer membrane (0.92).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 70A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot and for FACS analysis (Figure 70B).

The cp6270 protein was also identified in the 2D-PAGE experiment (Cpn0013).

These experiments show that cp6270 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 71

The following *C.pneumoniae* protein (PID 4376402) was expressed <SEQ ID 141; cp6402>:

1 MNVADLLSHL ETLSSKIFQ DYGPNGLQVG DPQTPVKKIA VAVTADLETI  
 51 KQAVAAEANV LIVHHGIFWK GMPYPITGMI HKRIQLLIEH NIQLIAYHLP  
 101 LDAHPTLGNN WRVALDLNWH DLKPFSSLP YLGVQGSFSP IDIDSFIDLL  
 151 SQYYQAPLKG SALGGPSRVS SAALISGGAY RELSSAATSQ VDCFITGNFD  
 201 EPAWSTALES NINFLAFGHT ATEKVGPKSL AEHLKSEFPI STTFIDTANP  
 251 F\*

The cp6402 nucleotide sequence <SEQ ID 142> is:

1 ATGAATGTTG CGGATCTCCT TTCTCATCTT GAGACTCTTC TCTCATCAAA  
 51 AATATTTTCT GATTATGGAC CCAACGGACT TCAAGTTGGA GATCCCCAAA  
 101 CTCCGGTAAA GAAAATCGCT GTTGCAGTTA CCGCAGATCT AGAAACCATA  
 151 AAACAAGCTG TTGCGGCCGA AGCAAACGTT CTCATTGTAC ACCACGGAAT  
 201 TTTTGGAAA GGTATGCCCT ATCCTATTAC CGGCATGATC CATAAGCGCA  
 251 TCCAATTACT AATAGAACAC AATATCCAAC TCATTGCCCTA CCACCTTCCT  
 301 TTGGATGCTC ACCCTACCTT AGGAAATAAC TGGAGAGTTG CCCTGGATCT  
 351 AAATTGGCAT GACTTGAAGC CCTTTGGTTC TTCCCTCCCT TATTTAGGAG  
 401 TGCAAGGCTC TTTCTCTCCT ATCGATATAG ATTCTTTCAT TGACCTGTTA  
 451 TCTCAATATT ACCAAGCTCC CCTAAAAGGA TCTGCCTTGG GCGGCCCTC  
 501 TAGAGTCTCC TCAGCAGCTC TGATCTCAGG AGGAGCTTAT AGAGAATCTT  
 551 CTTGCGCAGC CACGTCCCAA GTCGATTGCT TCATCACAGG AAATTTTGAT  
 601 GAACCTGCAT GGTGACAGC TCTAGAAAGC AATATCAACT TCCTAGCAAT  
 651 TGGACATACA GCCACAGAAA AAGTAGGTCC AAAATCTCTT GCAGAGCATC

5 651 TACAAGTTGG TTTACTGGAG CTGGACTCTA TCACCCAGAT ATTGTTGAAC  
 701 AAGATAGCTT GGCAATTACG AATTACCTAC ATAATAACGG GTACGCTGAT  
 751 GCTATAGTCA ACTCTCACTA TGACCTTGAC GACAAAGGGA ATATTCTTCT  
 801 TTACATGGAT ATTGATCGAG GGTCCGCGATA TACCTTAGGA CACGTCCATA  
 851 TCCAAGGGTT TGAGGTTTGT CCAAAACGCC TTATAGAAAA GCAATCCCAA  
 901 GTCGGCCCCA ATGATCTTTA TTGCCCCGAT AAAATATGGG ATGGGGCTCA  
 951 TAAGATCAAA CAAACTTATG CAAAGTATGG CTACATCAAT ACCAATGTAG  
 1001 ACGTTCTCTT CATCCCTCAC GCAACCCGCC CTATTTATGA TGTAACCTAT  
 1051 GAGGTAAGTG AAGGGTCTCC TTATAAAGTT GGGTTAATTA AAATTACTGG  
 1101 GAATACCCAT ACAAATCTG ACGTTATTTT ACACGAAACC AGTCTCTTCC  
 1151 CAGGAGATAC ATTCAATCGC TTAAAGCTAG AAGATACTGA GCAACGTTTA  
 1201 AGAATAACAG GCTACTTCCA AAGCGTAGT GTCTATACAG TTCGTTCTCA  
 1251 ACTTGATCCT ATGGGCAATG CGGATCAATA CCGAGATATT TTTGTAGAAG  
 1301 TCAAAGAAAC AACAACAGGA AACTTAGGCT TATTCCTAGG ATTTAGTTCT  
 1351 CTTGACAATC TTTTGGAGG AATTGAACTA TCTGAAAGTA ATTTTGATCT  
 1401 ATTTGGAGCT AGAAATATAT TTTCTAAAGG TTTTCGTGTG CTAAGAGGCG  
 1451 GTGGAGAACA TCTATCTTA AAAGCCAAC TCGGGGACAA AGTCACAGAC  
 1501 TATACTTTGA AGTGGACCAA ACCTCATTTT CTAAACACTC CTTGGATTTT  
 1551 AGGAATTGAA TTAGATAAAT CAATTAACAG AGCATTATCT AAAGATTATG  
 1601 CTGTCCAAAC CTATGGCGGG AACGTCAGCA CAACGTATAT CTTGAACGAA  
 1651 CACCTGAAAT ACGGTCTATT TTATCGAGGA AGTCAAACGA GTTTACATGA  
 1701 AAAACGTAAG TTCCTCCTAG GGCCAAATAT AGACAGCAAT AAAGGATTTG  
 1751 TCTCTGCTGC AGGTGTCAAC TTGAATTACG ATTCTGTAGA TAGTCCTAGA  
 1801 ACTCCAAC TA CAGGGATTCTG CCGGGGGGTG ACTTTTGAGG TTTCTGGTTT  
 1851 GGGAGGAAC TATCATTTTA CAAACTCTC TTAAACAGC TCTATCTATA  
 1901 GAAAACCTAC GCGTAAAGGT ATTTTGAAAA TCAAAGGGGA AGCTCAATTT  
 1951 ATTAAACCCT ATAGCAATAC TACAGCTGAA GGAGTTCCTG TCAGTGAGCG  
 2001 CTTCTTCTA GGTGGAGAGA CTACAGTTCG GGGATATAAA TCCTTTATTA  
 2051 TCGGTCCAAA ATACTCTGCT ACAGAACCTC AGGGAGGACT CTCTTCGCTC  
 2101 CTTATTTTCA AAGAGTTTCA ATACCCTCTC ATCAGACAAC CTAATATTAG  
 2151 TGCCTTTGTA TTCTTAGACT CAGGTTTGTG CGGTTTACAA GAGTATAAGA  
 2201 TTTCTGTTAA AGATCTACGT AGTAGTGCTG GATTTGGTCT GCGCTTCGAT  
 2251 GTAATGAATA ATGTTCTGT TATGTTAGGA TTTGGTTGGC CCTTCCGTCC  
 2301 AACCGAGACT TTGAATGGAG AAAAAATTGA TGTATCTCAG CGATTCTCT  
 2351 TTGCTTTAGG GGGCATGTTT TAA

The PSORT algorithm predicts outer membrane (0.7658).

The protein was expressed in *E.coli* and purified as GST-fusion (Figure 74A), his-tag and his-tag/GST-fusion products. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 74B) and for FACS analysis (Figure 74C).

40 The cp6576 protein was also identified in the 2D-PAGE experiment (Cpn0300).

These experiments show that cp6576 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 75

The following *C.pneumoniae* protein (PID 4376607) was expressed <SEQ ID 149; cp6607>:

45 1 MNKRQKDKLK ICVIISTLIL VGIFARAPRG DTFKTFKSE EAIYSNQCN  
 51 EDMRKILCDA IEHADEEIFL RIYNLSEPKI QQSLTRQAQA KNKVITYYQK  
 101 FKIPQILKQA SNVTLVEQPP AGRKLMHQKA LSIDKKDAWL GSANYTNLSL  
 151 RLDNNLILGM HSELCDLII TNTSGDFSII DQTGKYFVLP QDRKIAIQAV  
 201 LEKIQTAKT IQVMFALTH SEIIQALHQA KQGIHVDII IDRSHSKLTF  
 50 251 KQLRQLNINK DFVSINTAPC TLHHKFVID NKTLLAGSIN WSKGRFSLND  
 301 ESLIILENLT KQONQKLMI WKDLAKHSEH PTVDDEEKEI IEKSLPVEEQ  
 351 EAA\*

A predicted signal peptide is highlighted.

The cp6607 nucleotide sequence <SEQ ID 150> is:



201 PAS\*

The cp6567 nucleotide sequence <SEQ ID 146> is:

```

1  ATGACCTCAC CGATCCCCTT TCAGTCTAGT GGCGATGCCT CTTTCCTTGC
5  51  CGAGCAGCCA CAGCAACTCC CGTCTACTTC TGAATCTCAG CTAGTAAGTC
101 AATTGCTAAC CATGATGAAG CATACTCAAG CATTATCCGA AACGGTTCTT
151 CAACAACAAC GCGATCGATT ACCAACCACA TCTATTATCC TTCAAGTAGG
201 AGGAGCTCCT ACAGGAGGAG CGGGTGCGCC TTTTCAACCA GGACCGGCAG
251 ATGATCATCA TCATCCCATC CCGCCGCCTG TTGTACCAGC TCAAATAGAA
10 301 ACAGAAATCA CCACTATAAG ATCCGAGTTA CAGCTCATGC GATCTACTCT
351 ACAACAAAGC ACAAAGGAG CTCGTACAGG AGTTCTAGTG GTTACTGCAA
401 TCTTAATGAC GATCTCCTTA TTGGCTATTA TTATCATAAT ACTAGCTGTG
451 CTTGGATTTA CGGGCGTCTT GCCTCAAGTA GCTTTATTGA TGCAGGGTGA
501 AACAAATCTG ATTTGGGCTA TGGTGAGCGG TTCTATTATT TGCTTTATTG
15 551 CGCTAATTGG AACTCTAGGA TTAATTTTAA CAAATAAGAA CACGCCTCTA
601 CCGGCTTCTT AA

```

The PSORT algorithm predicts inner membrane (0.694).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 73A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 73B) and for FACS analysis.

20 These experiments show that cp6567 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 74

The following *C.pneumoniae* protein (PID 4376576) was expressed <SEQ ID 147; cp6576>:

```

25 1  MLIMRNKVIL QISILALIQT PLTLFST TEKV KEGHVVDSDI TIITEGENAS
51 51  NKHPLPKLKT RSGALFSQLD FDEDLRILAK EYDSVEPKVE FSEGKTNIAL
101 HLIAPKSIRN IHISGNQVVP EHKILKTLQI YRNDLFEREK FLKGLDDLRT
151 YYLKRGYFAS SVDYSLEHNQ EKGHIDVLIK INEGPCGKIK QLTFSGISRS
201 EKSDIQEFIQ TKQHSTTTSW FTGAGLYHPD IVEQDSLAIT NYLHNNGYAD
30 251 AIVNSHYDLD DKGNILLYMD IDRGSRYTLG HVHIQGFVL PKRLIEKQSQ
301 VGPNDLYCPD KIWDGAHKIK QTYAKYGYIN TNVDVLFIPH ATRPIYDVITY
351 EVSEGSYPYK GLIKITGNTH TKSDVILHET SLFPGDTFNR LKLEDTEQRL
401 RNTGYFQSVS VYTVRSQLDP MGNADQYRDI FVEVKETTTG NLGLFLGFSS
451 LDNLFGGIEL SESNFDLFGA RNIFSKGFRC LRGGGEHLFL KANFGDKVTD
501 YTLKWKPHF LNTPWILGIE LDKSINRALS KDYAVQTYGG NVSTTYILNE
35 551 HLKYGLFYRG SQTSLHEKRR FLLGPNIDSN KGFVSAAGVN LNYDSVDSPR
601 TPTTGIRGGV TFEVSGLGGT YHFTKLSLNS SIYRKLTRKG ILKIKGEAQF
651 IKPYSNNTAE GVPVSEFFL GGETTVRGYK SFIIGPKYSA TEPQGLLSSL
701 LISEEFQYPL IRQPNISAFV FLDSGFVGLQ EYKISLKDRL SSAGFGLRFD
751 VMNNVPVMLG FGWPFRPTET LNKEKIDVSQ RFFFALGGMF *

```

40 A predicted signal peptide is highlighted.

The cp6576 nucleotide sequence <SEQ ID 148> is:

```

1  ATGCTCATCA TGCAGAAATA AGTTATCTTG CAAATATCTA TTCTAGCGTT
51 51  AATCCAAACC CCTTTAACTT TATTTTCTAC TGAAAAAGTT AAAGAAGGCC
45 101 ATGTGGTGGT AGACTCTATC ACAATCATAA CGGAAGGAGA AAATGCTTCA
151 AATAAACATC CCTTACCCAA ATTAAGACC AGAAGTGGGG CTCCTTTTTC
201 TCAATTAGAT TTTGATGAAG ACTTGAGAA TCTAGCTAAA GAATACGACT
251 CTGTGTAGCC TAAAGTAGAA TTTTCTGAAG GAAAACTAA CATAGCCCTT
301 CACCTAATAG CTAAACCCTC AATTGGAAT ATTCTATCT CAGGAAATCA
351 AGTCGTTTCT GAACATAAAA TTCTTAAAC CCTACAAAT TACCGTAATG
50 401 ATCTCTTTGA ACGAGAAAAA TTTCTTAAGG GTCTTGATGA TCTAAGAACG
451 TATTATCTCA AGCGAGGATA TTTCGCATCC AGTGTAGACT ACAGTCTGGA
501 ACACAATCAA GAAAAAGGTC ACATCGATGT TTTAATTAAA ATCAATGAAG
551 GTCCTTGCGG GAAAAATAAA CAGCTTACGT TCTCAGGAAT CTCTCGATCA
601 GAAAAATCAG ATATCCAAGA ATTTATTCAA ACCAAGCAGC ACTCTACAAC

```

951 ATTAGGAGGG GTGGCTCTTG AATGTCAAAG ATGA

The PSORT algorithm predicts inner membrane (0.168).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 76A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 76B) and for FACS analysis.

The cp6624 protein was also identified in the 2D-PAGE experiment.

These experiments show that cp6624 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 77

10 The following *C.pneumoniae* protein (PID 4376728) was expressed <SEQ ID 153; cp6728>:

```

1  MKSSVSWLFF SSIPLFSSLS IVAAEVLDS SNNSYDGSNG TTFVSTTD
51  AAAGTYSLL SDVSFQNAGA LGIPLASGCF LEAGGDLTFQ GNQHALKFAF
101 INAGSSAGTV ASTSAADKNL LFNDFSRLSI ISCPSSLSP TGQCALKSVG
151 NLSLTGNSQI IFTQNFSSDN GGVINTKNFL LSGTSQFASF SRNQFTGKQ
15  201 GGVVYATGTI TIENSPGIVS FSQNLAKGSG GALYSTDNCS ITDNFQVIFD
251 GNSAWEEAAQ QGGAICCTTT DKTVTLTGNK NLSFTNMNTAL TYGGAISGLK
301 VSISAGGPTL FQSNISGSSA GQGGGGAINI ASAGELALSA TSGDITFNNN
351 QVTNGSTSTR NAINIIDTAK VTSIRAATGQ SIYFYDPITN PGTAASDTL
401 NLNLADANSE IEYGGAI VFS GEKLSPT EKA IAAVNTSTIR QPAVLARGDL
20  451 VLRDGVTVTF KDLTQSPGSR ILMGGTTL S AKEANLSLNG LAVNLSSLDG
501 TNKAALKTEA ADKNISLSGT IALIDTEGSF YENHNLKSAS TYPLLELTTA
551 GANGTITLGA LSTLTQEP E THYGYQGNWQ LSWANATSSK IGSINWRTG
601 YIPSPERKSN LPLNSLWGNF IDIRSINQLI ETKSSGEPFE RELWLSGIAN
651 FFYRDSMPTR HGFRHISGGY ALGITATTPA EDQLTF AFCQ LFARDRNHIT
25  701 GKNHGDYGA SLYFHHT EGL FDIANFLWGK ATRAPWVLSE ISQIIPLSFD
751 AKFSYLHTDN HMKTYT DNS I IKGSWRND A FCADLGASLP FVISVPYLLK
801 EVEPFVKVQY IYAHQQDFYE RHAEGRAF NK SELINVEIPI GVTFERDSKS
851 EKGT YDLTLM YILDAYRRNP KCQTS LIASD ANWMAYGTNL ARQGF SVRAA
901 NHFQVNP HME IFGQFAFEVR SSSRNYNTNL GSKFCF*
```

30 The cp6728 nucleotide sequence <SEQ ID 154> is:

```

1  ATGAAGTCCT CTGTCTCTTG GTTGTCTTTT TCTTCAATCC CGCTCTTTTC
51  ATCGCTCTCT ATAGTCGCGG CAGAGGTGAC CTTAGATAGC AGCAATAATA
101 GCTATGATGG ATCTAACGGA ACTACCTTCA CGGTCTTTTC CACTACGGAC
151 GCTGCTGCAG GAACTACCTA TTCCTTACTT TCCGACGTAT CCTTTCAAAA
35  201 TGCAGGGGCT TTAGGAATTC CTTAGCCTC AGGATGCTTC CTAGAAGCGG
251 GCGGCGATCT TACTTTCCAA GGAAATCAAC ATGCACTGAA GTTTCATTT
301 ATCAATGCGG GCTCTAGCGC TGGAAGTGTA GCCAGTACCT CAGCAGCAGA
351 TAAGAACTCT CTCTTTAATG ATTTTCTAG ACTCTCTATT ATCTCTGTCT
401 CCTCTCTTCT TCTCTCTCCT ACTGGACAAT GTGCTTTAAA ATCTGTGGGG
40  451 AATCTATCTC TAACTGGCAA TTCCCAAATT ATATTACTC AGAACTTCTC
501 GTCAGATAAC GCGGTGTTA TCAATACGAA AAAGTCTTCA TTATCAGGGA
551 CATCTCAGTT TGCGAGCTTT TCGAGAAACC AAGCCTTCAC AGGGAAGCAA
601 GCGGTGTAG TTTACGCTAC AGGAAGTATA ACTATCGAGA ACAGCCCTGG
651 GATAGTTTCC TTCTCTCAA ACCTAGCGAA AGGATCTGGC GGTGCTCTGT
45  701 ACAGCACTGA CAAGTGTTCG ATTACAGATA ACTTTCAGT GATCTTTGAC
751 GGCAATAGTG CTTGGGAAGC CGCTCAAGCT CAGGCGGGGG CTATTTGTTG
801 CACTACGACA GATAAAACAG TGACTCTTAC TGGGAACAAA AACCTCTCTT
851 TCACAAATAA TACAGCATTG ACATATGGCG GAGCCATCTC TGGACTCAAG
901 GTCAGTATTT CCGCTGGAGG TCCTACTCTA TTTCAAAGTA ATATCTCAGG
50  951 AAGTAGCGCC GGTACGGGAG GAGGAGGAGC GATCAATATA GCATCTGCTG
1001 GGGAACTCGC TCTCTCTGCT ACTTCTGGAG ATATTACCTT CAATAACAAC
1051 CAAGTCACCA ACGGAAGCAC AAGTACAAGA AACGCAATAA ATATCATTTGA
1101 TACCGCTAAA GTCACATCGA TACGAGCTGC TACGGGGCAA TCTATCTATT
1151 TCTATGATCC CATCACAAAT CCAGGAACCG CAGCTTCTAC CGACACATTG
55  1201 AACTTAAACT TAGCAGATGC GAACAGTGAG ATCGAGTATG GGGGTGCGAT
1251 TGTCTTTTCT GGAGAAAAGC TTTCCCTTAC AGAAAAGCA ATCGCTGCAA
```

-115-

1 ATGAATAAAA GACAAAAAGA TAAATTAAAA ATCTGTGTTA TTATTAGCAC  
 51 GTTGATTTTA GTAGGAATTT TTGCAAGAGC TCCTCGTGGT GACACTTTTA  
 101 AGACTTTTTT AAAGTCTGAA GAAGCTATCA TCTACTCAAA TCAATGCAAT  
 151 GAGGACATGC GTAAAATTCT ATGCGATGCT ATAGAACACG CTGATGAAGA  
 201 GATCTTCCTA CGTATTTATA ACCTCTCAGA ACCCAAGATC CAACAGAGTT  
 251 TAACTCGACA AGCTCAAGCA AAAAACAAAG TTACGATCTA CTATCAAAAA  
 301 TTTAAAATTC CCCAAATCTT AAAGCAAGCC AGCAATGTAA CTTTAGTCEGA  
 351 GCAACCTCCA GCAGGGCGTA AACTGATGCA TCAAAAAGCT CTTTCCATAG  
 401 ATAAGAAAGA TGCTTGGCTA GGATCTGCGA ACTACACCAA TCTTCTCTTA  
 451 CGTTTAGATA ATAATCTCAT TCTAGGAATG CATAGCTCGG AGCTCTGTGA  
 501 TCTCATTATC ACAAATACCT CTGGAGACTT TTCTATAAAG GATCAAAACAG  
 551 GAAAGTATTT TGTTCCTCCT CAAGATCGTA AAATTGCAAT ACAAGCTGTA  
 601 CTCGAAAAAA TCCAGACAGC TCAGAAAACC ATCCAAGTTG CTATGTTTGC  
 651 TCTGACCCAC TCGGAGATTA TTCAAGCCTT ACATCAAGCA AAACAACGAG  
 701 GAATCCATGT AGATATTATC ATTGATAGAA GTCATAGCAA ACTTACTTTT  
 751 AAGCAATTAC GACAATTAAA TATCAATAAA GACTTTGTTT CTATAAATAC  
 801 CGCACCTGT ACTCTTCACC ATAAGTTTGC AGTTATAGAT AATAAAACTC  
 851 TACTTGACAG ATCTATAAAT TGGTCTAAAG GAAGATTCTC CTTAAATGAT  
 901 GAAAGCTTGA TCATACTGGA AAACCTGACC AAACAACAAA ATCAGAAACT  
 951 TCGAATGATT TGGAAAGATC TAGCTAAGCA TTCAGAACAT CCTACAGTAG  
 1001 ACGATGAAGA AAAAGAAATT ATAGAAAAAA GTCTTCCAGT AGAAGAGCAA  
 1051 GAAGCAGCGT GA

The PSORT algorithm predicts periplasmic (0.934).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 75A) and also as a  
 25 GST-fusion. The GST-fusion protein was used to immunise mice, whose sera were used in a Western  
 blot (Figure 75B) and for FACS analysis.

These experiments show that cp6607 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 76

30 The following *C.pneumoniae* protein (PID 4376624) was expressed <SEQ ID 151; cp6624>:

1 MDAKMGYIFK VMRWIFCFVA CGITFGCTNS GFQANNSRPC ILSMNRMIHD  
 51 CVERVVG NRL ATAVLIKGS L DPHAYEMVKG DKDKIAGSAV IFCNGLGLEH  
 101 TSLSLRKHLEN NPNSVKLGER LIARGAFVPL EEDGICDPHI WMDLSIWKEA  
 151 VIEITEVLIE KFPEWSAEFK ANSEELVCEM SILD SWAKQC LSTIPENLRY  
 201 LVSGHNAFSY FTRRYLATPE EVASGAWRSR CISPEGLSPE AQISVRDIMA  
 251 VVDYINEHDV SVVFPEDTLN QDALKKIVSS LKKSHLVRLA QKPLYSDNVD  
 301 DNYFSTFKHN VCLITEELGG VALECQR\*

The cp6624 nucleotide sequence <SEQ ID 152> is:

1 ATGGATGCGA AAATGGGATA TATATTTAAA GTGATGCGTT GGATTTTCTG  
 40 51 TTTCGTGGCA TGTGGTATAA CTTTGGATG TACCAATTCT GGGTTTCAGA  
 101 ATGCAAATTC ACGTCCTTGT ATACTATCCA TGAATCGCAT GATTCATGAT  
 151 TGTGTTGAAA GAGTCGTGGG GAATAGGCTT GCTACCGCTG TTTTGATCAA  
 201 AGGATCCCTTA GACCTCATG CGTATGAGAT GGTAAAGGG GATAAGGACA  
 251 AGATTGCTGG AAGTCCGTA ATTTTGTGTA ACGGCCTGGG TCTTGAGCAT  
 45 301 ACATTAAGTT TCGGGAAGCA TTTAGAAAAT AATCCCAATA GTGTCAAGTT  
 351 AGGGGAGCGG TTGATAGCGC GTGGGGCCTT TGTTCCTCTA GAAGAAGACG  
 401 GTATTTGCGA TCCTCATATC TGGATGGATC TTTCTATTG GAAGGAAGCT  
 451 GTCATAGAAA TTACAGAAGT TCTCATTGAA AAGTTCCCTG AATGGTCTGC  
 501 TGAATTTAAA GCAAATAGTG AGGAAC TTGTGAAATG TCTATTTTAG  
 551 ATTCTTGGGC GAAACAATGC TTGAGCACAA TTCCTGAAA TTTACGGTAT  
 601 CTTGCTCTAG GTCATAATGC GTTCAGTTAC TTTACACGTC GCTATTTAGC  
 651 TACTCTGAA GAAGTGGCTT CCGGAGCATG GAGGTCTCGT TGTATTTCTC  
 701 CTGAGGGTCT ATCTCCAGAA GCTCAAATCA GTGTTGCTGA TATTATGGCG  
 751 GTTGTAGATT ATATTAATGA GCATGATGTC AGTGTGGTTT TCCCTGAGGA  
 801 TACTCTGAAC CAAGATGCGT TGAAAAAAT TGTTCCTCT CTGAAGAAAA  
 851 GTCATTTAGT TCGTCTAGCT CAAAACCAT TGTATAGTGA TAATGTGGAC  
 901 GACAATTATT TTAGCACCTT TAAACATAAT GTCTGCCTTA TCACAGAAGA

-118-

```

1  ATGTTTCGTAA  TGAAAAAACT  TGTCCGCTCTA  TGCCTAGTTC  TTCTTTCTTT
51  ACTTCCGAAT  GTATTATTTT  CTTCCGATCT  TTTACGAGAA  GAGGGCATCA
101  AAAAGATGAT  GGACAAGCTG  ATCGAGTATC  ATGTCGATGC  TCAAGAGGTT
151  TCTACGGATA  TACTCTCGCG  TTCTTTATCT  AGTTACATTC  AATCTTTTGA
5  201  TCCTCATAAA  TCTTATCTTT  CAAACCAAGA  GGTTCAGTT  TTTCTACAGT
251  CTCCGGAAC  AAAGAAACGT  CTCTTAAAGA  ATTATAAGGC  AGGCAACTTT
301  GCTATTTATC  GCAACATCAA  TCAATTAAAT  CATGAGAGTA  TTCTTCGTGC
351  CAGGCAGTGG  AGAAACGAAT  GGGTTAAGAA  TCCAAAAGAG  CTTGTATTGG
10  401  AGGCATCCTC  ATATCAGATA  TCGAAGCAAC  CTATGCAATG  GAGCAAATCT
451  TTAGACGAAG  TGAAGCAGAG  ACAACGCGCT  CTAATCCTTT  CCTATCTTTC
501  TTTACATCTT  GCTGGAGCTT  CTTCTCTCTG  TTATGAGGGT  AAAGAAGAGC
551  AGCTTGCTGC  TCTGTGCTA  CGTCAAATCG  AGAACCATGA  GAATGTATAT
601  TTAGGTATCA  ACGATCATGG  TGTTGCTATG  GATCGGGATG  AAGAAGCCTA
651  CCAATTCCAT  ATCCGTGTTG  TTAAAGCTTT  AGCTCATAGC  TTAGATGCAC
15  701  ATACGGCGTA  TTTCAGTAAG  GACGAAGCGT  TGGCGATGCG  AATCCAACTA
751  GAAAAAGGCA  TGTGTGGAAT  TGGTGTGTT  CTGAAGGAAG  ATATTGATGG
801  AGTTGTGTT  AGAGAAATCA  TTCCTGGGG  ACCTGCGGCT  AAATCTGGGG
851  ATCTTCAGCT  TGGAGATATC  ATCTATCGGG  TGGATGGCAA  GGATATCGAG
901  CATCTTTCTT  TCCGCGGTGT  TTTAGATTGT  TTACGTGGAG  GTCATGGCTC
20  951  TACTGTAGTC  TTAGATATCC  ATCGTGGGGA  GAGCGATCAT  ACGATCGCCT
1001  TGAGAAGGGA  GAAAATCCTT  TTAGAAGACC  GTCGTGTGGA  TGTTTCCTAT
1051  GAGCCTTATG  GAGATGGTGT  GATTGGGAAA  GTTACGTTAC  ATTCTTTTAA
1101  TGAAGGAGAA  AATCAGGTTT  CTAGTGAACA  AGATCTACGT  CGAGCGATTG
1151  AGGGATTAAA  GGAGAAGAAC  CTTCTTGGAT  TAGTTTTAGA  TATCCGAGAA
25  1201  AATACGGGTG  GATTTTTATC  TCAAGCGATC  AAAGTTTCTG  GTTTATTTAT
1251  GACCAATGGC  GTTGTGGTTG  TATCTCGCTA  TGCTGATGGT  ACCATGAAGT
1301  GCTACCGCAC  AGTATCTCCT  AAAAAATTCT  ATGATGGTCC  TTTGGCTATT
1351  TTAGTATCTA  AAAGTTCCGC  ATCAGCAGCG  GAGATTGTAG  CACAACTCT
30  1401  CCAAGATTAT  GGAGTTGCTT  TAGTTGTTGG  AGATGAGCAG  ACCTATGGGA
1451  AGGGAACGAT  TCAGCATCAA  ACAATTACTG  GAGATGCCCT  TCAGGACGAT
1501  TGTTTTAAGG  TTACTGTAGG  GAAATATTAT  TCCCCTTCTG  GGAAATCGAC
1551  TCAACTTCAG  GGAGTAAAAT  CCGATATTTT  AATTCCTTCT  CTCTATGCTG
1601  AAGATCGTCT  AGGAGAGCGT  TTTCTAGAGC  ATCCCTTACC  TGCAGATTGC
1651  TGTGATAATG  TACTTCACGA  TCCTCTCACG  GACTTGGATA  CTCAAACACG
35  1701  TCCTTGTTTT  CAAAAATACT  ATCTTCCTAA  TCTACAAAAG  CAAGAGACTC
1751  TTTGGAGAGA  GATGCTACCT  CAGCTTACGA  AAAACAGTGA  GCAAAGGCTT
1801  TCTGAGAATT  CGAATTTTCA  GGCATTTTTC  TCGCAGATAA  AATCATCTGA
1851  AAAAACGGAC  CTATCCTATG  GTTCCAATGA  TTTACAATTG  GAAGAGTCGA
1901  TAAACATTTT  GAAGGACATG  ATTTTATTAC  AACAGTGTAG  AAAATAA

```

40 The PSORT algorithm predicts periplasmic (0.932).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 78A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 78B) and for FACS analysis.

45 These experiments show that cp6847 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 79

The following *C.pneumoniae* protein (PID 4376969) was expressed <SEQ ID 157; cp6969>:

```

1  MRLFSLGITY LFFSLALSSC CGYSILNSPY HLSSLGKSL L QERIFIAPIK
51  EDPHGQLCSA LTYELSKRSF AISGRSSCAG YTLKVELLNG IDKNIGFTYA
50  101  PNKLGDKTHR HFIVSNEGRL SLSAKVQLIN NDTQEVLLIDQ CVARESVDFFD
151  FEPDLGTANA HEFALGQFEM HSEAIKSARR ILSIRLAETI AQQVYYDLF*

```

A predicted signal peptide is highlighted.

The cp6969 nucleotide sequence <SEQ ID 158> is:

```

55  1  ATGAGATTGT  TTTCTTTAGG  CACGATTTAT  CTTTTTTTTT  CTCTAGCACT
51  51  TTCGTCATGC  TGTTGGTTACT  CTATTTTAAA  CAGCCCGTAT  CACTTATCGT
101  101  CTTTAGGTAA  GTCTTTATTA  CAGGAAAGAA  TTTTCATTGC  TCCCATAAAA

```

-117-

5 1301 ACGTCACCTC TACTATCCGA CAACCTGCAG TATTAGCGCG GGGAGATCTT  
 1351 GTACTTCGTG ATGGAGTCAC CGTAACTTTC AAGGATCTGA CTCAAAAGTCC  
 1401 AGGATCCCGC ATCTTAATGG ATGGGGGGAC TACACTTAGT GCTAAAGAGG  
 1451 CAAATCTTTC GCTTAATGGC TTAGCAGTAA ATCTCTCCTC TTTAGATGGA  
 1501 ACCAACAAGG CAGCTTTAAA AACAGAAGCT GCAGATAAAA ATATCAGCCT  
 1551 ATCGGGAACG ATTGCGCTTA TTGACACGGA AGGGTCATTTC TATGAGAATC  
 1601 ATAACCTTAAA AAGTGCTAGT ACCTATCCTC TTCTTGAAC TACCACCGCA  
 1651 GGAGCCAACG GAACGATTAC TCTGGGAGCT CTTTCTACCC TGACTCTTCA  
 10 1701 AGAACCTGAA ACCCACTACG GGTATCAAGG AAACCTGGCAG TTGTCTTGGG  
 1751 CAAATGCAAC ATCCTCAAAA ATAGGAAGCA TCAACTGGAC CCGTACAGGA  
 1801 TACATTCTTA GTCCTGAGAG AAAAAGTAAT CTCCCTCTAA ATAGCTTATG  
 1851 GGGAACTTT ATAGATATAC GCTCGATCAA TCAGCTTATA GAAACCAAGT  
 1901 CCAGTGGGGA GCCTTTTGAG CGTGAGCTAT GGCTTTCAGG AATTGCGAAT  
 1951 TTCTTCTATA GAGATTCTAT GCCCACC CGC CATGGTTTCC GCCATATCAG  
 15 2001 CGGGGGTTAT GCACTAGGGA TCACAGCAAC AACTCCTGCC GAGGATCAGC  
 2051 TTACTTTTGC CTTCTGCCAG CTCCTTGCTA GAGATCGCAA TCATATTACA  
 2101 GGTAAAGAAC ACGGAGATAC TTACGGTGCC TCTTTGTATT TCCACCATAC  
 2151 AGAAGGGCTC TTCGACATCG CCAATTTCCT CTGGGGAAAA GCAACCCGAG  
 2201 CTCCCTGGGT GCTCTCTGAG ATCTCCAGA TCATTCCTTT ATCGTTCGAT  
 20 2251 GCTAAATTCA GTTATCTCCA TACAGACAAC CACATGAAGA CATATTATAC  
 2301 CGATAACTCT ATCATCAAGG GTTCTTGGAG AAACGATGCC TTCTGTGCAG  
 2351 ATCTTGGAGC TAGCCTGCCT TTTGTTATTT CCGTTCGGTA TCTTCTGAAA  
 2401 GAAGTCGAAC CTTTTGTCAA AGTACAGTAT ATCTATGCGC ATCAGCAAGA  
 2451 CTTCTACGAG CGTCATGCTG AAGGACGCGC TTTCAATAAA AGCGAGCTTA  
 25 2501 TCAACGTAGA GATTCCTATA GGCGTCACCT TCGAAAGAGA CTCAAATCA  
 2551 GAAAAGGGAA CTTACGATCT TACTCTTATG TATATACTCG ATGCTTACCG  
 2601 ACGCAATCCT AAATGTCAAA CTTCCCTAAT AGCTAGCGAT GCTAACTGGA  
 2651 TGGCCTATGG TACCAACCTC GCACGACAAG GTTTTCTGT TCGTGCTGCG  
 2701 AACCATTTC AAGTGAACCC CCACATGGAA ATCTTCGGTC AATTCGCTTT  
 30 2751 TGAAGTACGA AGTTCCTTAC GAAATTATAA TACAAACCTA GGCTCTAAGT  
 2801 TTTGTTTCTA G

The PSORT algorithm predicts inner membrane (0.187).

35 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 77A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 77B) and for FACS analysis.

The cp6728 protein was also identified in the 2D-PAGE experiment.

These experiments show that cp6728 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 78

40 The following *C.pneumoniae* protein (PID 4376847) was expressed <SEQ ID 155; cp6847>:

45 1 MFVMKKLVRL CVVLLSLLPN VLFSSDLLRE EGIKKMMDKL IEYHVDAQEV  
 51 STDILSRSL S YIQSFDPHK SYLSNQEVAV FLOSPETKKR LLKNYKAGNF  
 101 AIYRNINQLI HESILRARQW RNEWVKNPKE LVLEASSYQI SKQPMQWSKS  
 151 LDEVKQRQRA LLLSYLSLHL AGASSRYEG KEEQLAALCL RQIENHENVY  
 201 LGINDHGVAM DRDEEAYQFH IRVVKALAHS LDAHTAYFSK DEALAMRIQL  
 251 EKGMCIGIVV LKEDIDGVVV REIIPGGPAA KSGDLQLGDI IYRVDGKDIE  
 301 HLSFRGVLD LRGHGHSTVV LDIHRGESDH TIALRREKIL LEDRRVDVSY  
 351 EPYGDGVIGK VTLHSFYEGE NQVSSEQDLR RAIQGLKEKN LLGLVLDIR  
 401 NTGGFLSQAI KVSGLFMTNG VVVVSRYADG TMKCYRTVSP KKFYDGPLAI  
 50 451 LVSKSSASAA EIVAQTLQDY GVALVVGDEQ TYGKGTIQHQ TITGDASQDD  
 501 CFKVTVGKYY SPSGKSTQLQ GVKSDILIPS LYAEDRLGER FLEHPLPADC  
 551 CDNVLHDPLT DLDQTQRPWF QKYYLPNLQK QETLWREMLP QLTKNSEQRL  
 601 SENSNFQAF L SQIKSSEKTD LSYGSNDLQL EESINILKDM ILLQQCRK\*

A predicted signal peptide is highlighted.

55 The cp6847 nucleotide sequence <SEQ ID 156> is:

These experiments show that cp7109 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 81

The following *C.pneumoniae* protein (PID 4377110) was expressed <SEQ ID 161; cp7110>:

```

5      1  MAAIKQILRS MLSQSSLWMV LFSLYSLSGY CYVITDKPED DPHSSSAVKW
      51  DHWGKTTLRS LSNKKASAKA VSGTGATTVG FIKDTWSRTY AVRWNWYWGK
     101  ELPTSSWVKK SKATGISSDG SIIAGIVENE LSQSFAVTWK NNEMYLLPST
     151  WAVQSKAYGI SSDGSVIVGS AKDAWSRTFA VKWTGHEAQV LPVGWAVKSV
     201  ANSVSANGSI IVGSVQDASG ILYAVKWEKN TITHLGTLLG YSAIAKAVSN
10    251  NGKVIVGRSE TTYGVEVHAF C HKNGVMSDLG TLGGSYSAAK GVSATGKVIV
     301  GMSTTANGKL HAFKYVGGRM IDLGEYSWKE ACANAVSIDG EIIVGVQSE*

```

A predicted signal peptide is highlighted.

The cp7110 nucleotide sequence <SEQ ID 162> is:

```

15      1  ATGGCAGCTA TAAACAAAT TTTACGTTCT ATGCTATCTC AGAGTAGCTT
      51  ATGGATGGTC CTATTTTCAT TATATCTCTC ATCTGGTTAT TGCTATGTAA
     101  TTACAGACAA ACCAGAAGAT GACTTCCATT CTTTCATCCGC AGTAAATGG
     151  GATCATTGGG GAAAGACAAC TCTCTCAAGA TTATCAAATA AAAAAGCCTC
     201  TGCAAAAGCT GTTTCAGGAA CTGGTGCTAC AACTGTCGGC TTTATAAAAG
     251  ACACCTGGTC TCGAACATAC GCAGTAAGAT GGAATTATTG GGGGACCAAA
20    301  GAACTCCCTA CCAGCTCATG GGTAAAAAAA TCAAAAGCAA CAGGAATCTC
     351  CTCTGATGGG TCTATAATCG CGGGGATTGT CGAGAATGAG CTTTCTCAAA
     401  GTTTCGCAGT CACATGGAAA AACAATGAAA TGTATTGCT CCCTTCCACA
     451  TGGGCAGTGC AATCTAAAGC GTATGGAATT TCTTCTGATG GCTCTGTTAT
     501  TGTAGGGAGT GCTAAGGATG CTTGGTCGCG AACTTTCGCT GTGAAGTGGA
25    551  CGGGACACGA GGCTCAGGTG TTACCAGTAG GCTGGGCTGT CAAATCTGTA
     601  GCGAATTCTG TACTTGCCAA TGGATCTATA ATTGTAGGGT CTGTACAAGA
     651  CGCCTCTGGA ATTCTTTATG CTGTAAAGTG GGAAGGGAAC ACTATTACAC
     701  ATCTAGGAAC TTTAGGAGGC TATTCTGCCA TTGCAAAGC TGTATCCAAT
     751  AATGGCAAGG TCATTGTAGG GAGATCCGAA ACATATTATG GAGAGTCCA
30    801  TGCTTCTGT CATAAGAATG GCGTCATGTC AGACCTCGGC ACCCTCGGAG
     851  GATCTTATTC TGCAGCTAAG GGAGTCTCTG CAACTGGAAA AGTTATTGTC
     901  GGTATGTCCA CAACAGCAA TGGGAAATTG CATGCCTTGA AATATGTCGG
     951  TGGAAGAATG ATCGACTTAG GAGAGTATAG CTGGAAAGAA GCCTGTGCAA
1001  ACGCTGTTTC TATTGATGGA GAAATTATTG TTGGAGTCCA ATCAGAATAA

```

35 The PSORT algorithm predicts outer membrane (0.827).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 81A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 81B) and for FACS analysis.

40 These experiments show that cp7110 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Figure 191 shows a schematic representation of the structural relationships between of cp7105, cp7106, cp7107, cp7108, cp7109 and cp7110, each of which is identified herein. These six proteins may be grouped in a new family of related outer membrane-associated proteins. These proteins have a repeat structure in common (*cf.* the *pmp* family).

### 45 Example 82

The following *C.pneumoniae* protein (PID 4377127) was expressed <SEQ ID 163; cp7127>:

```

1  MVFFRNSLLH LVALSGMLCC SSGVALTIAE KMASLEHSGR GADDYEGMAS

```

```

151 GAAGATCCTC ATGGTCAGCT CTGCTCAGCT CTAACCTATG AGCTTAGTAA
201 GCGTTCTTTT GCTATCTCTG GAAGGAGTTC TTGCGCAGGC TATACTCTTA
251 AAGTAGAGCT TCTGAATGGT ATTGACAAGA ATATAGGTTT TACGTATGCC
301 CCAAATAAAC TCGGAGATAA GACTCACAGG CATTTTATAG TCTCTAATGA
351 AGGCAGACTA TCACTATCTG CAAAAGTACA GCTTATCAAT AATGACACTC
401 AAGAAGTCCT TATAGACCAA TGTGTTGCTC GAGAGTCTGT AGACTTTGAC
451 TTTGAGCCTG ACTTAGGAAC AGCAAACGCT CATGAATTG CTTTAGGCCA
501 ATTTGAAATG CATAGTGAAG CCATAAAAAG TGCTCGCCGT ATACTATCTA
551 TACGCCTAGC CGAGACGATT GCTCAACAGG TATACTATGA CCTTTTTTGA

```

10 The PSORT algorithm predicts inner membrane (0.126).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 79A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 79B) and for FACS analysis.

15 These experiments show that cp6969 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 80

The following *C.pneumoniae* protein (PID 4377109) was expressed <SEQ ID 159; cp7109>:

```

1 MKKTCCQNYR SIGVVFSVVL FVLTTQTLFA GHFIDIGTSG LYSWARGVSG
51 DGRVVVGYEG GNAFKYVDGE KFLLEGLVPR SEALVFKASY DGSVIIGISD
101 QDPSCRAVKW VNGALVDLGI FSEGMQSFAE GVSSDGKTIV GCLYSDDET
151 NFAVKWDETG MVVLPNLPED RHSCAWDASE DGSVIVGDAM GSSEIAKAVY
201 WKDGEQHLLS NIPGAKRSSA HAVSKDGSFI VGEFISEENE VHAFVYHNGV
251 IKDITGLGGD YSVATGVSRD GKVIVGHSTR TDGEYRAFKY VDGMRIDLGT
301 LGGSASFAGF VSDDGKTIVG KFETELGECH AFIYLLDD*

```

25 A predicted signal peptide is highlighted.

The cp7109 nucleotide sequence <SEQ ID 160> is:

```

1 ATGAAAAAGA CATGTTGCCA AAATTACAGA TCGATAGGCG TTGTGTTCTC
51 TGTGGTACTT TTCGTTCTTA CAACACAGAC GCTGTTTGCA GGACATTTTA
101 TTGATATTGG AACTTCTGGA TTATATTCTT GGGCTCGAGG TGTATCTGGA
30 151 GATGGCCGCG TTGTCTGAGG TTATGAAGGT GGCAATGCAT TTAAATATGT
201 TGATGGTGAG AAATTTCTGT TAGAAGGTTT GGTCCCGAGA TCCGAGGCCT
251 TGGTATTTAA AGCTTCTTAT GATGGCTCTG TAATTATAGG AATCTCGGAT
301 CAAGATCCGT CTTGCCGCGC TGTGAAGTGG GTAAACGGTG CACTTGTTGA
35 351 TCTTGGAATA TTTTCTGAGG GAATGCAATC TTTTGCAGAG GGTGTTTCCA
401 GTGATGGAAA GACGATTGTA GGTGCCTAT ATAGTGATGA TACAGAGACA
451 AACTTTGCTG TGAAGTGGGA TGAAACAGGA ATGGTTGTTT TCCCTAACTT
501 ACCAGAAGAT CGACATTCTT GCGCTTGGA TGCCTCTGAA GATGGCTCTG
551 TGATTGTAGG GGACGCCATG GGTAGCGAGG AAATTGCCAA GGCAGTGTA
601 TGGAAGGACG GTGAACAACA TCTGCTTTCT AATATCCCAG GAGCTAAAAG
40 651 ATCGTCAGCA CATGCAGTTT CTAAAGATGG ATCTTTTATC GTAGGCGAGT
701 TCATCAGTGA AGAAAATGAA GTTCATGCCT TTGTTTATCA CAACGGTGTT
751 ATCAAAGATA TCGGGACTTT AGGAGGAGAT TACTCTGTAG CAACTGGAGT
801 TTCTAGGGAT GGTAAAGTCA TCGTGGGTCA TTCTACAAGA ACAGATGGTG
851 AATACCGTGC ATTTAAATAT GTGGATGGAA GAATGATAGA TTTGGGGACT
45 901 TTAGGAGGTT CAGCATCTTT TGCTTTTGGT GTTCTGACG ATGGCAAAAC
951 AATCGTAGGA AAATTGAAA CAGAGCTAGG AGAATGTCAT GCCTTTATCT
1001 ACCTTGATGA TTAG

```

The PSORT algorithm predicts outer membrane (0.887).

50 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 80A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 80B) and for FACS analysis.

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 82A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 82B) and for FACS analysis.

These experiments show that cp7127 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 83

The following *C.pneumoniae* protein (PID 4377133) was expressed <SEQ ID 165; cp7133>:

```

1  MQPFIFTLLC LTSLVSLVAF DAANARKRCA CAQTIERGEN FFSIKRSACA
51  EIEYQEKSRH ASAIERISKD KGKVTPKQIA KVATKKKQRY RLLQVPFSRP
101 PNNSRYNLYA LLSEPPECYS DTASWYAIFI RLLRRAYVDT GNVPPGSEYA
151 IANALISNKQ EILERGAQLG PDVIETLTLP EEQAEIFYKM LKGSSNSQSL
201 LNFLHYEEKS LGHCKLNLIF MDPLLLEAVL DHPDAYRETS LLRDGIWEAV
251 KRQEHAIQEH GQAAALELFK TRTDFRLELR DKMQLLLSRY DLLPLLNKKM
301 FDYTLGSAGD YLFLVDPDTK AISRCRCPSK SIKL

```

A predicted signal peptide is highlighted.

The cp7133 nucleotide sequence <SEQ ID 166> is:

```

1  ATGCAACCTT TTATCTTTAC TTTACTGTGC TTGACATCTT TGGTTTCTTT
51  AGTCGCCTTT GATGCTGCGA ATGCTCGTAA ACGTTGTGCC TGTGCTCAAA
101 CTATAGAACG TGGAGAGAAC TTCTTTTCCA TAAAACGCTC TGCTTGTGCT
151 GAAATCGAAT ATCAAGAAAA ATCTCGCCAC GCCTCAGCAA TTGAAAGAAT
201 CTCAAAAGAT AAAGGCAAAG TCACTCCAAA GCAGATTGCG AAAGTAGCTA
251 CTAAGAAAAA GCAAAGATAC CGTTTATTGC AGGTTCCTT TTCAAAGCCT
301 CCGAATAACT CAAGGTATAA CCTCTATGCT TTGCTTAGTG AACCTCCCGA
351 ATGCTATAGC GATACAGCAT CATGGTATGC TATTTTTTAT CGGTTACTTC
401 GACGTGCTTA TGTAGACACG GGAAATGTAC CTCCTGGATC TGAGTATGCC
451 ATCGCTAATG CTTTGATAAG TAACAAACAA GAGATTTTAG AGAGGGGAGC
501 CCGAGCTTGA CCCGATGTTA TTGAAACTCT AACATTGCCT GAGGAACAAG
551 CGAGATTTT TTATAAAATG CTCAAAGGGT CGTCAAACTC TCAGTCGCTA
601 CTGAATTTTC TGCATTATGA AGAGAAAAGC TTAGGCCACT GTAAGCTAAA
651 TCTGATCTTC ATGGATCCCC TACTGTTAGA AGCTGTTCTA GATCATCCCG
701 ATGCTTATAG GGAAACGTCG CTCCTGCGCG ATGGCATTTG GGAAGCGGTG
751 AAGCGTCAAG AACATGCCAT CCAAGAACAT GGCCAGGCAG CTGCTTTGGA
801 GCTTTTTTAAA ACACGCACCG ACTTCCGCCT GGAGCTGCGA GATAAGATGC
851 AGTTACTTCT AAGTCGATAC GATTTGCTCC CCTTATTAAA TAAAAAATG
901 TTCGACTACA CCTTAGGAAG TGCCGGAGAT TACTTATTTT TGGTAGACCC
951 AGATACTAAG GCAATTTCTC GATGTCGCTG CCCTTCAAAG AGTATTAAAT
1001 TATAA

```

The PSORT algorithm predicts outer membrane (0.92).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 83A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 83B) and for FACS analysis.

These experiments show that cp7133 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 84

The following *C.pneumoniae* protein (PID 4377222) was expressed <SEQ ID 167; cp7222>:

```

1  MNRRDMVITA VVVNAILLVA LFVTSKRIGV KDYDEGFRNF ASSKVTOAVV
51  SEEKVIEKPV VAEVPSRPIA KETLAAQFIE SKPVIVTPP VPVVSETPEV

```



-121-

51 FNANMREYSL QLSKLYEEAR KLRASGTEDE ALWKDLIRRI GEVRGYLREI  
 101 EELWAAEIRE KGGNLEDYAL WNHPEITTYN LVTDYGTEDS IYLIPQEIGA  
 151 IKIATLSKFV VPKESEFEDCL TQILSRLGIG VRQVNSWIKE LYMMRKEGCS  
 201 VAGVFSSRKD LEALPETAAYI GFVLNSNVDA HTNQHVLLKF INPETHVDV  
 5 251 IAGRVWIFGS AGEVGELLKI YNFVQSESIR QEYRVIPLTK IDPGEMISIL  
 301 NAAFREDLTK DVSEESLGLR VVPLQYQGRS LFLSGTAALV QQALTFLIREL  
 351 EEGIENPTDK TVFWYNVKHS DPQELAALLS QVHDFVSGEN KASVGAADGC  
 401 GSQNLASIQI DTTVSSSAKD GSVKYGNFIA DSKTGTLMV VEKEVLPRIQ  
 451 MLLKKLDVPK KMRVIEVLLF ERKLAHEQKS GLNLLRLGEE VCKKGCSPSV  
 10 501 SWAGGTGILE FLFKGSTGSS IVPGYDLAYQ FLMAQEDVRI NASPSVVTMN  
 551 QTPARIAVVD EMSIAVSSDK DKAQYNRAQY GIMIKMLPVI NVGEEDGKSY  
 601 ITLETDTITFD TTGKNHDDRP DVTRRNITNK VRIADGETVI IGGRLCKQMS  
 651 DSHDGIPFLG DIPGIGKLFQ MSSTSDSLTE MFVFITPKIL ENPVEQQRK  
 701 EEALLSSRPG EREYYQALA ASEAAARAAH KKLEMPFASG VSLSQVERQE  
 15 751 YDGC\*

A predicted signal peptide is highlighted.

The cp7127 nucleotide sequence <SEQ ID 164> is:

1 ATGGTTT TTT TCCGTAATTC TTTACTGCAT TTAGTTGCCC TATCCGGAAT  
 51 GCTCTGTTGT TCTTCTGGAG TGGCTTTAAC GATAGCCGAG AAGATGGCTT  
 20 101 CTTTAGAGCA CTCGGGGAGA GGAGCAGACG ATTATGAGGG GATGGCTTCG  
 151 TTTAATGCCA ATATGAGGGA GTATAGCCTT CAGCTGAGCA AGTTGTATGA  
 201 GGAAGCACGA AAGCTACGCG CTTCTGGAAC TGAGGATGAA GCTCTGTGGA  
 251 AGGACTTAAT TCGACGGATT GGTGAGGTGC GAGGCTATCT TCGAGAGATC  
 301 GAGGAGCTTT GGGCTGCAGA AATTCGTGAG AAAGGGGGCA ATCTCGAGGA  
 25 351 CTACGCCCTC TGAATCACCC CAGAGACTAC GATTTACAAT CTTGTTACCG  
 401 ATTACGGAAC CGAAGACTCT ATTTATTTGA TTCCTCAAGA AATCGGAGCG  
 451 ATTAAAATCG CAACCTTATC GAAATTTGTA GTTCCTAAAG AGTCTTTTCA  
 501 AGACTGTCTC ACTCAGATCC TATCTCGCTT AGGTATTGGC GTGCGTCAGG  
 551 TCAATCTTGT GATTAAAGAA CTTTATATGA TCGTAAGGA GGGCTGCAGT  
 30 601 GTTGCTGGAG TTTTTCCTC CAGAAAAGAT TTAGAGGCGC TCCCGAAAC  
 651 AGCCTATATT GGTTTTGTAT TGAATTCGAA CGTAGATGCG CATACCAATC  
 701 AACATGTCTT AAAAAAGTTC ATTAACCCTG AAACAACGCA TGTAGATGTG  
 751 ATTGCAGGAC GTGTGTGGAT TTTTGGTTCT GCGGGGGAAG TCGGCGAGCT  
 801 TCTGAAGATT TATAATTTTG TGCAGTCGGA GAGCATACGT CAAGAGTATC  
 35 851 GGGTGATTC CTTAACTAAG ATCGATCCAG GGGAGATGAT TTCCATTCTC  
 901 AACCGAGCAT TTCGTGAGGA TCTGACTAAA GATGTTAGTG AAGAATCTTT  
 951 AGGCCTTCGT GTAGTTCCCT TACAGTATCA AGGGCGTTCG TTGTTTTTAA  
 1001 GTGGAACCGC GGCCTTAGTG CAGCAAGCGC TGAATCTCAT TCGAGAGCTT  
 1051 GAAGAAGGGA TTGAGAACCC TACGGATAAA ACAGTATTTT GGTATAACGT  
 40 1101 CAAGCACTCC GATCCCCAAG AGTTGGCGGC ATTGCTTTCC CAAGTCCATG  
 1151 ATGCTCTTCT TGGCGAGAAT AAGGCGAGTG TCGGAGCTGC AGATGGATGT  
 1201 GGGTCGCAAT TAAATGCCTC GATCCAAAT TACTACTACG TAAGTTCTTC  
 1251 TGCGAAAGAT GGCTCAGTGA AGTACGGAAA CTTCATCGCG GATTCTAAGA  
 1301 CAGGAACCTCT GATTATGTTG GTTGAGAAAG AAGTTCTTCC ACGTATTACG  
 45 1351 ATGCTACTTA AGAACTAGA TGTCCCTAAA AAGATGGTCC GTATCGAGGT  
 1401 GCTGTTATTT GAAAGAAAAT TGGCACATGA GCAGAAATCT GGGTTAAATC  
 1451 TTCTACGCTT TGGTGAGGAA GTTTGTAAAA AAGGGTGCAG TCCTTCTGTG  
 1501 TCTTGGGCCG GGGGTACTGG CATACTAGAA TTTTATTTA AAGGAAGTAC  
 1551 GGGACTTTTC ATAGTTCCCT GTTATGATCT CGCCTATCAA TTTTAAATGG  
 50 1601 CTCAAGAGGA CGTTCGGATT AATGCGAGTC CTTCTGTAGT TACTATGAAC  
 1651 CAAACCCAG CACGGATTGC TGTGTTGAT GAAATGTCAA TAGCGGTGTC  
 1701 TTCAGATAAA GATAAGCGC AATACAATCG TCGCAGTAC GGTATCATGA  
 1751 TAAAAATGCT CCCCCTAATT AATGTGGGAG AGGAAGACGG AAAAAGTTAC  
 1801 ATTACTTTAG AGACAGACAT CACCTTTGAT ACTACGGGAA AAAATCATGA  
 55 1851 TGATCGTCCT GATGTTACAA GCGTAATAT TACTAATAAG GTGCGCATTG  
 1901 CTGACGGAGA GACTGTGATT ATTGGAGGTT TGCCTTGCAA ACAGATGTCA  
 1951 GATTCTCATG ATGGCATTC TTTCTTTGGA GACATTCTTG GTATAGGGAA  
 2001 GTTATTTGGA ATGAGTTCCA CATCAGACAG TCTCAGGAG ATGTTTGTAT  
 2051 TTATCACTCC GAAGATCCTA GAAAATCCTG TAGAGCAACA AGAACGTAAA  
 60 2101 GAAGAAGCTT TACTCTCTTC GCGCCCTGGA GAGAGAGAAG AATACTATCA  
 2151 GGCTTTAGCA GCTAGTAGAG CTGCAGCAG AGCAGCTCAT AAAAAATTAG  
 2201 AGATGTTCCC GGCATCAGGA GTATCTTTAT CTCAGGTAGA GAGGCAAGAA  
 2251 TACGATGGCT GCTAG

The PSORT algorithm predicts periplasmic (0.920).

-124-

401 CCTATGCTAT TGGAGGACTC GCTGCAAACT GCCTGAATGG GTATTCTGGA  
 451 TCATCGAAAA TCTTCGTTGC CGAAGCCGAT GAAAGTGATG GGTCTTTAAA  
 501 GCACTACACT CCCCCTGCAG TAGTCATTAC AAATATAGAT AATGAACATT  
 551 TGAATAATTA CGCTGGGAAT CTTGATAACC TGGTTCAGGT AATCCAGGAC  
 601 TTCTCTAGAA AAGTAACAGA TCTCAATAAG GTATTCTATA ACGGGGATTG  
 651 TCCTATTTTG AAAGGAAATG TCCAAGGGAT TTCCTATGGA TATTCACCAG  
 701 AATGTCAATT GCATATCGTT TCCTATAATC AAAAGGCATG GCAATCTCAC  
 751 TTTTCCTTTA CCTTTTTAGG CCAGGAGTAT CAAGACATTG AGCTCAATCT  
 801 CCCTGGACAA CATAACGCTG CAAATGCAGC AGCAGCCTGT GGAGTTGCTC  
 851 TTACCTTTTG CATAGACATA AACATCATTG GAAAAGCTCT CAAAAAATTC  
 901 TCGGGAGTTC ATCGACGCTC AGAAAGAAAA AATATATCCG AAAGCTTTCT  
 951 TTTCTTAGAA GATTATGCTC ATCATCCTGT AGAGGTGCA CATACCCTGC  
 1001 GCTCTGTGCG TGATGCTGTG GGTTCGCGAA GAGTCATCGC AATTTTTCAA  
 1051 CCACATCGAT TCTCTCGTTT AGAAGAGTGC TTACAAACCT TCCCCAAAGC  
 1101 TTTCCAAGAA GCTGATGAAG TCATACTTAC AGATGTCTAT AGTGCCGGAG  
 1151 AAAGTCCTAG AGAGTCTATC ATTCTTTCCG ACCTTGCGGA ACAGATTCCG  
 1201 AAGTCTTCTT ATGTCCATTG TTGTATGTT CCCCATGGAG ACATCGTAGA  
 1251 TTATCTACGA AACTACATTG GCATTATGA TGTCTGTGTT TCTCTAGGAG  
 1301 CTGGAAATAT CTATACTATT GGAGAGGCTT TAAAAGACTT TAACCCTAAA  
 1351 AAATTATCCA TAGGACTCGT CTGTGGAGGG AAATCTTGGC AACACGATAT  
 1401 TTCTCTACTT TCTGCTCAAC ATGTCTCTAA ATATATTCTT CCTGAATTTCT  
 1451 ATGATGTGAG TTAATTCATC ATAAATCGTC AGGGCTTATG GAGAACAGGA  
 1501 AAGGATTTTC CTCATCTTAT TGAAGAGACT CAAGGGGATT CGCCACTTTC  
 1551 TTCTGAAATC GCTTCAGCTT TAGCAAAAGT CGACTGTTTG TTTCCCGTGC  
 1601 TCCATGGCCC ATTTGGAGAG GATGGTACGA TCCAGGGATT TTTTGAAATC  
 1651 TTAGGAAAAC CTTATGCCGG ACCCTCACTA TCTTTAGCAG CAACTGCAAT  
 1701 GGATAAGCTG TTAACAAAAC GAATTGCATC AGCAGTGGGT GTTCCCTGTAG  
 1751 TCCCTTACCA ACCTTTAAAT CTCTGTTTCT GGAAACGCAA TCCAGAACTA  
 1801 TGTATTCAGA ATCTTATAGA GACATTTTCT TTCCCTATGA TTGTAAAAAC  
 1851 TGCACATTTG GGATCTAGTA TTGGGATATT TTTAGTCCGT GATAAAGAGG  
 1901 AATTACAAGA AAAGATCTCA GAAGCATTTC TATATGACAC GGATGTGTTT  
 1951 GTGGAGGAAA GTCGCTTAGG GTCTCGTGAA ATCGAAGTGT CCTGTATCGG  
 2001 CCATTCTTCT AGCTGGTATT GTATGGCAGG GCCTAATGAA CGCTGTGGTG  
 2051 CTAGTGGGTT TATTGATTAT CAAGAGAAAT ATGGATTGTA TGGCATAGAT  
 2101 TGCACAAAGA TCTCTTTTGA TTTACAGCTC TCACAAGAAAT CTTTAGATTG  
 2151 TGTTAGAGAA CTTGCAGAGC GTGTCTACCG AGCAATGCAA GGAAAAGGTT  
 2201 CAGCTCGAAT AGATTTTTTC TTGGATGAAG AGGGGAATTA TTGGTTGTCA  
 2251 GAGGTCAATC CTATTCCAGG AATGACAGCA GCTAGCCCAT TTTTACAAGC  
 2301 TTTTGTTTAC GCAGGATGGA CGCAAGAACA AATTGTAGAT CACTTTATTA  
 2351 TAGATGCTCT ACATAAGTTT GATAAGCAGC AGACTATCGA ACAGGCATTG  
 2401 ACTAAAGAAC AAGATTTAGT TAAAAGATAA

The PSORT algorithm predicts inner membrane (0.16).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 85A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 85B) and for FACS analysis.

These experiments show that cp7225 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 86

The following *C.pneumoniae* protein (PID 4377248) was expressed <SEQ ID 171; cp7248>:

50 1 MKFWLQGCFA VGCLLLTLPC CAARRRASGE NLQQTRPIAA ANLQWESYAE  
 51 ALEHSKQDHK PICLFFTGSD WCMWCIKMD QILQSSEFKH FAGVHLHMVE  
 101 VDFPQKNHQP EEQRQKNQEL KAQYKVTGFP ELVFIDAEGK QLARMGFEPG  
 151 GGAAYVSKVK SALKLR\*

A predicted signal peptide is highlighted.

55 The cp7248 nucleotide sequence <SEQ ID 172> is:

1 ATGAAATTTT GGTGCAAGG ATGTGCTTTT GTCGGTTGTC TGCTATTGAC

-123-

101 PTVAVPPQPV RETVKEEQAP YATVVVKKGD FLERIRANH TTVAKLMQIN  
 151 DLTTTQLKIG QVIKVPTSQD VSNEKTPQTQ TANPENYYIV QEGDSPWTIA  
 201 LRNHIRLDDL LKMNLDLEYK ARRLKPGDQL RIR\*

A predicted signal peptide is highlighted.

5 The cp7222 nucleotide sequence <SEQ ID 168> is:

1 ATGAATCGTA GAGACATGGT AATAACAGCT GTCGTAGTGA ATGCTATATT  
 51 GCTTGTGGCT CTTTTCGTCA CATCAAAGCG TATTGGCGTC AAGGACTATG  
 101 ACGAGGGATT CCGTAATTTT GCTTCTAGCA AGGTTACACA AGCAGTAGTT  
 151 TCAGAAGAAA AAGTCATAGA AAAGCCTGTA GTCGCAGAAG TGCCTAGCCG  
 10 TCCTATCGCT AAAGAGACTC TAGCTGCACA GTTTATTGAA AGTAAGCCGG  
 251 TTATTGTAAC CACACCACCC GTGCCTGTTG TTAGCGAAAC CCCAGAAGTG  
 301 CCTACTGTGG CAGTTCCGCC TCAGCCTGTT CGTGAGACAG TAAAAGAGGA  
 351 ACAAGCTCCT TATGCTACTG TTGTAGTGAA AAAAGGAGAT TTTCTCGAAC  
 401 GCATTGCGAG AGCAAATCAT ACTACCGTTG CAAAATTGAT GCAGATCAAT  
 15 GATCTTACCA CCACCCAAC TAAAATTGGT CAGGTCATCA AAGTCCCTAC  
 501 GTCTCAAGAT GTCAGCAACG AAAAAACTCC TCAAACACAG ACCGCAAACC  
 551 CTGAAAATTA TTATATCGTC CAAGAAGGGG ATAGCCCGTG GACAATAGCA  
 601 TTGCGTAACC ATATTGATG GGATGATTG CTAAAAATGA ATGATCTCGA  
 651 TGAATATAAA GCCCGGCGCC TTAAGCCTGG AGATCAGTTG CGCATACGTT  
 20 701 GA

The PSORT algorithm predicts periplasmic (0.935).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 84A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 84B) and for FACS analysis.

25 These experiments show that cp7222 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 85

The following *C.pneumoniae* protein (PID 4377225) was expressed <SEQ ID 169; cp7225>:

1 MKGTPQYHFI GIGGIGMSAL AHILLDRGYE VSGSDLYESY TIESLKAKGA  
 30 51 RCFSGHDSH VPHDAVVVYS SSIAPDNVEY LTAIQRSSRL LHRAELLSQL  
 101 MEGYESILVS GSHGKTGTSS LIRATFQEAQ KDPSYAIGGL AANCLNGYSG  
 151 SSKIFVAEAD ESDGSLKHYT PRAVVITNID NEHLNNYAGN LDNLVQVIQD  
 201 FSRKVTDLNK VFYNGDCPIL KGNVQGISYG YSPECQLHIV SYNQKAWQSH  
 251 FSFTFLGQEY QDIELNLPGQ HNAANAAAAC GVALTFGIDI NIIRKALKKF  
 35 301 SGVHRRLERK NISESFLFLE DYAHHPVEVA HTLRSVRDAV GLRRVIAIFQ  
 351 PHRFSLREEC LQTFPKAFQE ADEVILTDVY SAGESPRESI ILSDLAEQIR  
 401 KSSVHCCYV PHGDIVDYLR NYIRIHDVCV SLGAGNIYTI GEALKDFNPK  
 451 KLSIGLVCGG KSCEHDISLL SAQHVSKYIS PEFYDVSYFI INRQGLWRTG  
 501 KDFPHLIEET QGDSPLSSEI ASALAKVDCL FVVLHGPFGE DGTIQGFFEI  
 40 551 LGKPYAGPSL SLAATAMDKL LTKRIASAVG VPVVPYQPLN LCFWKRNPEL  
 601 CIQNLIETFS FPMIVKTAHL GSSIGIFLVR DKEELQEKIS EAFLYDTDVF  
 651 VESRLGSRE IEVSCIGHSS SWYCMAGPNE RCGASGFIDY QEKYGFDDID  
 701 CAKISFDLQL SQESLDCVRE LAERVYRAMQ GKGSARIDFF LDEEGNYWLS  
 751 EVNPIPGMTA ASPFLQAFVH AGWTQEQIVD HFII DALHKF DKQQTIEQAF  
 45 801 TKEQDLVKR\*

The cp7225 nucleotide sequence <SEQ ID 170> is:

1 ATGAAGGGAA CTCCTCAGTA TCATTTTATC GGTATCGGTG GTATAGGAAT  
 51 GAGCGCTTTA GCTCATATTT TGCTTGATCG TGGCTATGAG GTCTCTGGAA  
 101 GCGACTTATA TGAAAGCTAT ACGATCGAAA GCCTGAAAGC TAAAGGTGCG  
 151 AGGTGTTTCT CAGGCCATGA TTCCTCCCAT GTTCCTCATG ATGCCGTCGT  
 201 TGTTTATAGC TCAAGTATAG CCCCTGATAA TGTAAGTAT CTTACCGCTA  
 251 TTCAAAGATC ATCACGTCTT CTTCATAGAG CAGAGCTCTT GAGTCAGCTT  
 301 ATGGAGGGTT ATGAAAGCAT TCTGGTTTCA GGAAGCCATG GGAAGACAGG  
 351 GACCTCATCT CTAATTCGAG CGATTTTCCA GGAAGCTCAG AAAGATCCCT

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 87A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 87B) and for FACS analysis.

These experiments show that cp7249 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 88

The following *C.pneumoniae* protein (PID 4377261) was expressed <SEQ ID 175; cp7261>:

```

1  MLPISILLFY VILGCLSAIY ADKKKRNVIW WFFAGAFFGF IGLVLLLLLP
51  SRRNALEKPO NDPFDNSDLF DDLKKSILAGN DEIPSSGDLQ EIVIDTEKWF
101 YLNKDRENVG PISFEELVVL LKGKTYPEEI WVKKGKMDW QRVKDVPSLQ
151 QALKEASK*
```

The cp7261 nucleotide sequence <SEQ ID 176> is:

```

1  ATGCTCCCTA TTTCGATTTT ATTATTTTAT GTGATTCTAG GTTGTCTATC
51  TGCCTACATA GCAGATAAGA AAAAACGAAA TGTATTGGC TGGTTTTTTG
101 CAGGAGCATT TTTTGGATTG ATTGGTCTAG TTGTCCTTCT TCTTCTTCCT
151 TCTCGTCGAA ACGCTTTAGA AAAGCCACAA AACGATCCTT TTGATAACTC
201 CGATCTTTTT GATGATTTGA AAAAAAGTTT AGCAGGTAAT GACGAGATAC
251 CCTCATCGGG AGATCTTCAA GAAATCGTTA TCGATACAGA GAAGTGGTTT
301 TATTTAAATA AAGATAGAGA AAACGTAGGT CCGATATCTT TTGAGGAGTT
351 GGTCTGACTT TTAAAGGGAA AAACGTATCC AGAAGAAATT TGGGTATGGA
401 AAAAGGGAAT GAAAGATTGG CAACGAGTGA AGGATGTTCC ATCACTACAA
451 CAGGCTTTGA AAGAAGCATC AAAATAA
```

The PSORT algorithm predicts inner membrane (0.848).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 88A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 88B) and for FACS analysis.

These experiments show that cp7261 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 89

The following *C.pneumoniae* protein (PID 4377305) was expressed <SEQ ID 177; cp7305>:

```

1  MEVYSFHPAV RTSFQHRVMA ALDAWFFLGG HRLKVVSLDS CNSGWAYQEL
51  VSISTTEKVL KLLSYLLVPI VIALLIRCL LHSNFRIDVE KERWLKIREL
101 GIDIESCKLP SSYVNQVSSF IWFEKDKSKR PRIDVDYHTL HSKDWVVFPI
151 VFQKIPKTSR FSYWFSQKET RKRQYVRNML DHVIGYLTSE GGEWLQYISK
201 TSYQSATSLD PERVLQYCLT DNQELQGEVQ RLLNEESATK SSGDKEVLLS
251 HVSDIICQW WPKFLEVIQS PAFIEELVEE VSGKLNLDLFL CLEKANTLDQ
301 ELRNSLLRAV VHHGSEGVDI KKVAGLIYY TEAIQLQIPF SRS*
```

The cp7305 nucleotide sequence <SEQ ID 178> is:

```

1  ATGGAAGTTT ATAGTTTTCA CCCTGCGGTA AGGACTTCGT TTCAGCACCG
51  TGTAATGGCA GCACTAGATG CTTGGTTTTT TCTAGGAGGG CACCGTTTAA
101 AAGTAGTTTC TCTAGATAGT TGTAAGTCAG GTTGGGCGTA TCAAGAACTT
151 GTGTCTATTT CAACGACAGA AAAAGTCTTG AACTACTCTT CTTACCTACT
201 CGTACCGATT GTCATAATAG CTCTGTTAAT TCGTTGTCTT TTACATAGCA
251 ATTTTAGGAT AGACGTAGAG AAGGAACGTT GGTAAAAAAT AAGGGAGTTA
301 GGAATTGATA TAGAAAGCTG CAAACTCCCC AGTTCTTATG TAAACCAGGT
351 TTCCTCGTTT ATTTGGTTTG AAAAAGATAA ATCCAAACGG CCACGTATTG
401 ATGTAGATTA TCATACGCTA CATAGCAAAG ACTGGGTAGT TTTCCCTATC
```

-125-

51 TTTACCTTGT TGTGCTGCAC GAAGACGTGC TTCTGGAGAA AATTTGCAAC  
 101 AAACCTCGTCC TATAGCAGCT GCAAATCTAC AATGGGAGAG CTATGCAGAA  
 151 GCTCTTGAAC ATTCTAAACA AGATCACAAA CCTATTTGTC TTTTCTTTAC  
 201 AGGATCAGAC TGGTGTATGT GGTGCATAAA AATGCAAGAC CAGATTTTGC  
 5 251 AAAGCTCTGA GTTTAAGCAT TTTGCGGGTG TGCATCTGCA TATGGTTGAA  
 301 GTTGATTTC CCCAAAAGAA TCATCAACCT GAAGAGCAGC GCCAAAAAAA  
 351 TCAAGAACTG AAAGCTCAAT ATAAAGTTAC AGGATTCCCC GAACTGGTCT  
 401 TCATAGATGC AGAAGGAAAA CAGCTTGCTC GCATGGGATT TGAGCCTGGT  
 10 451 GGTGGAGCTG CTTACGTAAG CAAGGTGAAG TCTGCTCTTA AACTACGTTA  
 501 A

The PSORT algorithm predicts periplasmic (0.932).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 86A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 86B) and for FACS analysis.

15 The cp7248 protein was also identified in the 2D-PAGE experiment.

These experiments show that cp7248 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 87

The following *C.pneumoniae* protein (PID 4377249) was expressed <SEQ ID 173; cp7249>:

20 1 MIPSPTPINF RDDTILETDP KPSLIMFSSK KTEIASERRK AHPTLFKVLG  
 51 TIWNIVKFII SIILFLPLAL LWVLKKTCQF FILPSSIIISQ SMSKTAVAIR  
 101 RMTFLSHIKQ LLSLKEISAA DRVVIQYDDL VVDSLAIKIP HALPHRWILY  
 151 SQGNSGLMEN LFDGRDSSLH QLAATGSNL LVFNYPGIMS SKGEAKRENL  
 201 VKSYQACVRY LRDEETGPKA NQIIAFGYSL GTSVQAAALD REVTDGSDGT  
 25 251 SWIVVKDRGP RSLADVANI CKPIASAIK LVGWNIDSVK PSERLRCPEI  
 301 FIYNSNHDQE LISDGLFERE NCVATPFLEL PEVKTSGTKI PIPERDLLHL  
 351 NPLSPNVVDR LAAVISNYLD SENRKSQQPD \*

The cp7249 nucleotide sequence <SEQ ID 174> is:

30 1 ATGATCCCAT CCCCTACCCC AATAAACTTT CGTGATGATA CGATTCTAGA  
 51 GACGGATCCA AAGCCGTCTT TAATCATGTT CTCTTCAAAA AAAACAGAGA  
 101 TAGCTTCTGA AAGACGGAAG GCCCATCCCA CCTTATTTAA AGTTCCTAGGA  
 151 ACGATTGGA ATATTGTGAA GTTTATTATC TCAATCATTG TGTTCCTTCC  
 201 CTAGCGTTA TTGTGGGTAC TCAAGAAAAC CTGTCAGTTT TTCATTCTCC  
 25 251 CATCTTCTAT CATATCTCAG AGCATGTCAA AAACAGCTGT GGCAATTTCGG  
 301 CGAATGACCT TTCTGTCCCA TATTAAACAA CTCCTAAGCC TTAAGGAAAT  
 35 351 CTCAGCTGCC GATCGTGTGG TTATACAATA TGACGATTTG GTGGTTGATA  
 401 GCTTAGCTAT AAAGATACCT CATGCTCTTC CCCACAGGTG GATTCTTTAT  
 451 TCTCAAGGAA ACTCTGGATT GATGGAAAAC CTGTTTCGATC GGGGCGATTG  
 501 CTCTCTACAC CAGCTAGCCA AAGCAACCGG CTGGAATCTT CTGTGTGTTCA  
 40 551 ACTATCCTGG AATTATGTCC AGCAAAGGAG AAGCGAAACG AGAAAATCTG  
 601 GTTAAATCGT ATCAGGCATG CGTACGCTAC CTACGAGATG AAGAGACAGG  
 651 TCCTAAAGCC AATCAAATCA TAGCTTTCGG ATACTCTTTG GGAAGTAGTG  
 701 TCCAAGCTGC TGCTCTAGAT CGTGAGGTCA CTGATGGCAG TGATGGAAGT  
 751 TCATGGATTG TTGTAAAAGA TCGGGGCCCT CGCTCTCTAG CAGATGTGCG  
 45 801 GAATCAAATT TGTAAGCCCA TAGCTTCCGC GATTATAAAA CTCGTGTTGTT  
 851 GGAACATAGA CTCTGTGAAA CCTAGCGAAA GATTGCGTTG TCCCGAAATT  
 901 TTCAATTACA ACTCTAATCA TGATCAAGAA CTCATTAGCG ACGGCCTCTT  
 951 CGAAAGAGAA AATTGCGTAG CAACACCTTT TCTAGAGCTT CCTGAAGTAA  
 1001 AAACCTCGGG GACTAAATT CCTATACCCG AAAGGGATCT TCTCCATCTA  
 50 1051 AATCCTCTCA GTCCAAATGT AGTAGACAGA TTAGCAGCAG TGATCTCTAA  
 1101 TTATTTAGAT TCTGAAAACA GAAAGTCTCA GCAACCTGAT TAA

The PSORT algorithm predicts inner membrane (0.571).

1051 CATTGGA AAA AAGAGACTGA TGCTTTGATT ATTGATCAGA CCCATAATCC  
 1101 TGGAGGCAGT GTTTTCTATC TCTATTCGTT ACTATCTATG TTAACAGATC  
 1151 ATCCTTTAGA TACTCCTAAA CATAGAATGA TTTTCACTCA GGATGAAGTC  
 1201 AGCTCGGCTT TGCACGGA AGATCTACTA GAAGATGTCT TCACAGATGA  
 5 1251 GCAGGCAGTT GCCGTGCTAG GGGAACTAT GGAAGGATAT TGCATGGATA  
 1301 TGCATGCTGT AGCCTCTCTT CAAAACCTCT CTCAGAGTGT CCTTCTTCC  
 1351 TGGGTTTCAG GTGATATTAA CCTTCAAAA CCTATGCCTT TGCTAGGATT  
 1401 TGCACAGGTT CGACCTCATC CTAACATCA ATATACTAAA CCTTTGTTTA  
 1451 TGTTGATAGA CGAGGATGAC TTCTCTGTG GAGATTTAGC GCCTGCAATT  
 10 1501 TTGAAGGATA ATGGCCGCGC TACTCTCATT GGAAAGCCAA CAGCAGGAGC  
 1551 TGGAGGTTTT GTATTCCAAG TCACTTTCCC TAACCGTTCT GGAATTAAAG  
 1601 GTCTTTCTTT AACAGGATCT TTAGCTGTTA GGAAAGATGG TGAGTTTATT  
 1651 GAAACTTAG GAGTGGCTCC TCATATTGAT TTAGGATTTA CCTCCAGGGA  
 1701 TTTGCAAACT TCCAGGTTTA CTGATTACGT TGAGGCAGTG AAAACTATAG  
 15 1751 TTTTAACTTC TTTGTCTGAG AACGCTAAGA AGAGTGAAGA GCAGACTTCT  
 1801 CCGCAAGAGA CGCCTGAAGT TATTCGAGTC TCTTATCCCA CAACGACTTC  
 1851 TGCTTCGTAA

The PSORT algorithm predicts periplasmic space (0.2497).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 90A) and also in  
 20 his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a  
 Western blot (Figure 90B) and for FACS analysis.

These experiments show that cp7347 is a surface-exposed and immunoaccessible protein, and that it  
 is a useful immunogen. These properties are not evident from the sequence alone.

### Example 91

25 The following *C.pneumoniae* protein (PID 4377353) was expressed <SEQ ID 181; cp7353>:

1 MNMPVPSAVP SANITLKEDS STVSTASGIL KTATGEVLVS CTALEGSST  
 51 DALISLALGQ IILATQCELL LQSTNVHQLL FLPPEVVELE IQVVDLLVQL  
 101 EHAETITSEP QETQTQSRSE QTLPQQSSSK QSALSPRSLK PEISDSKQQQ  
 151 ALQTPKDSAV RKHSEAPSPE TQARASLSQA SSSSQRLPP QESAPERTLL  
 30 201 EQQKASSFSP LSQFSAEKQK EALTTSKSHE LYKERDQDRQ QREQHDRKHD  
 251 QEEDAESKKK KKKRGLGVEA VAEPEGNLD IAALIFSDQM RPPAETSCK  
 301 ETTFKKKLPS PMSVFSRFIP SKNPLSVGSS IHGPIQTPKV ENVFLRFMKL  
 351 MARILGQAEA EANELYMRVK QRTDDVDTLT VLISKINNEK KDIDWSENEE  
 401 MKALLNRAKE IGVITDKKEY TWTEEEKRLL KENVQMRKEN MEKITQMERT  
 35 451 DMQRHLQEIS QCHQARSNVL KLLKELMDTF IYNLRP\*

The cp7353 nucleotide sequence <SEQ ID 182> is:

1 ATGAATATGC CTGTTCTTTC TGCAGTTCCT TCTGCAAATA TAACTCTAAA  
 51 AGAAGACAGC TCAACAGTTT CCACAGCCTC TGGAATATTA AAGACTGCAA  
 101 CAGGTGAAGT CTTAGTCTCT TGTACAGCGC TAGAAGGAAG CTCTTCTACA  
 40 151 GATGCTTTAA TTAGCTTAGC TTTAGGACAA ATCATTCTTG CGACCCAACA  
 201 AGAACTGCTC TTACAAAGCA CAAATGTTCA TCAACTCCTC TTCCTCCCTC  
 251 CTGAAGTTGT AGAATTAGAA ATCCAAGTTG TTGACTTGCT AGTGCAATTG  
 301 GAACATGCAG AGACAATCAC AAGTGAACCA CAAGAAACAC AAACGCAAG  
 351 TAGGAGTGAG CAGACCCTCC CTCAACAAAG CAGCAGTAAA CAATCTGCTC  
 45 401 TCTCCCCACG CTCCTTAAAA CCTGAAATTT CTGATTCTAA ACAACAGCAA  
 451 GCTCTTCAAA CACCAAAAGA CTCTGCTGTA AGAAAACACA GCGAAGCACC  
 501 GTCACCTGAG ACACAAGCTC GCGCTTCCTT ATCTCAGGCA AGCTCAAGTT  
 551 CTCAGAGATC CTTACCTCCG CAAGAAAGTG CGCCAGAAAG AACACTATTA  
 601 GAACAACAAA AAGCAAGCTC CTTCTCTCCT CTATCCCAGT TCTCTGCAGA  
 50 651 GAAACAAAAA GAGGCCCTGA CGACCTCAA ATCTCATGAA CTCTATAAAG  
 701 AACGCGATCA AGATCGCCAA CAAAGAGAGC AGCACGACAG AAAGCACGAT  
 751 CAGGAAGAAG ACGCTGAATC TAAAAAGAAA AAGAAGAAAC GTGGTCTCGG  
 801 TGTAAGAGCA GTCGCTGAGG AACCCGGAGA AAATCTAGAT ATTGCCGCTT  
 851 TAATCTTCTC AGATCAAATG CGACCTCCTG CTGAAGAAAC TTCTAAAAAA  
 55 901 GAAACGACAT TCAAAAAGAA GCTACCTTCT CCAATGTCTG TGTTTAGCAG  
 951 ATTCAATCCCT AGTAAGAATC CGTTATCTGT AGGCTCTTCA ATACACGGGC  
 1001 CTATACAAAC TCCAAAAGTA GAAAATGTGT TCTTAAGGTT CATGAAGCTC

-127-

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55

```

451 GTTTTTCAGA AAATTCCAA GACCTCGCGT TTCAGTTATT GGTTCACACA
501 AAAAGAAACA AGGAAGAGGG ATTATGTGAG AAATATGCTG GACCACGTCA
551 TTGGTTATCT AACGTCAGAA GGTGGGGAGT GGTTCAGTA TATATCGAAA
601 ACCTCTTATC AAAGCGCTAC TTCCTTGGAT CCTGAAAGAG TTCTTCAATA
651 TTGCTTAACT GATAACCAGG AGCTCCAGGG AGAAGTGCAA CGTTTGCTTA
701 ATGAGGAGAG TGCAGACCAA AGCTCTGGGG ATAAGGAAGT TTTGTTAAGT
751 CATGTATCTG ACATTATTTG CCAGTGTGG TGGCCAAAGT TTCTTGAAGT
801 TATACAATCT CCGGCCCTTA TTGAAGAATT AGTAGAAGAA GTGAGTGGTA
851 AACTTAATTT AGATTTTSTA TGCCTAGAAA AGGCTAATAC ATTAGATCAG
901 GAGTTGAGAA ACAGTCTTCT AAGAGCAGTC GTACACCACG GTTCTGAAGG
951 AGTTGATATT AAGAAAGTTG GTGCCGCCCT CATTTATTTAT ACGGAAGCTA
1001 TTCAATTACA GATTCCTTC TCAAGGAGTT AA

```

The PSORT algorithm predicts inner membrane (0.508).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 89A) and also as a double GST/his fusion. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 89B) and for FACS analysis.

These experiments show that cp7305 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 90

20 The following *C.pneumoniae* protein (PID 4377347) was expressed <SEQ ID 179; cp7347>:

25  
30  
35  
40  
45  
50  
55

```

1 MKKGKLGATV FGLLFTSSVA GFSKDLTKDN AYQDLNVIEH LISLKYAPLP
51 WKELLFQWDL SQQTQARLQ LVLEEKPTTN YCQKVLNRYV RSLNDYHAGI
101 TFYRTESAYI PYVLKLSLEDG HVFVVDVQTS QGDIYLGDEI LEVDGMGIRE
151 AIESLRFGRG SATDYSAAVR SLTSRSAAFG DAVPSGIAML KLRRPSGLIR
201 STPVWRWYTP EHIGDFSLVA PLIPEHKPQL PTQSCVLFPS GVNSQSSSSS
251 LFSSYMPVYF WEELRVQNKQ RFDSNHHIGS RNFGLPTFGP ILWEQDKGPY
301 RSYIFKAKDS QGNPHRIGFL RISSYVWTDL EGLEEDHKDS PWELFGEIID
351 HLEKETDALI IDQTHNPGGS VFYLYSLLSM LTDHPLDTPK HRMIFTQDEV
401 SSALHWQDLL EDVFTDEQAV AVLGETMEGY CMDMHAVASL QNFSQSVLSS
451 WVSGDINLSK PMPLLGFAQV RPHPKHQYTK PLFMLIDEDD FSCGDLAPAI
501 LKDNNGRATLI GKPTAGAGGF VFQVTFPNRS GIKGLSLTGS LAVRKDGEFI
551 ENLGVAPHID LGFTSRDLQT SRFTDYVEAV KTIVLTSLSLSE NAKKSEEQTS
601 PQETPEVIRV SYPTTTSAS*

```

A predicted signal peptide is highlighted.

35 The cp7347 nucleotide sequence <SEQ ID 180> is:

40  
45  
50  
55

```

1 ATGAAAAAAG GGAAATTAGG AGCCATAGTT TTTGGCCTTC TATTTACAAG
51 TAGTGTGCT GGTTTTTCTA AGGATTTGAC TAAAGACAAC GCTTATCAAG
101 ATTTAAATGT CATAGAGCAT TTAATATCGT TAAAATATGC TCCTTTACCA
151 TGAAGGAAC TATTATTTGG TTGGGATTTA TCTCAGCAA CACAGCAAGC
201 TCGCTTGCAA CTGGTCTTAG AAGAAAAACC AACAACCAAC TACTGCCAGA
251 AGGTACTCTC TAACTACGTG AGATCATTAAC ACATTATCA TGCAGGGATT
301 ACGTTTTATC GTACTGAAAG TCGTATATC CCTTACGTAT TGAAGTTAAG
351 TGAAGATGGT CATGTCTTTG TAGTCGACGT ACAGACTAGC CAAGGGGATA
401 TTTACTTAGG GGATGAAATC CTTGAAGTAG ATGGAATGGG GATTCGTGAG
451 GCTATCGAAA GCCTTCGCTT TGGACGAGGG AGTGCCACAG ACTATTCTGC
501 TGCAGTTCGT TCCTTGACAT CGCGTTCGCG CGCTTTTGGA GATGCGGTTT
551 CTTCAGGAAT TGCCATGTTG AAACCTCGCC GACCCAGTGG TTTGATCCGT
601 TCGACACCGG TCCGTTGGCG TTATACTCCA GAGCATATCG GAGATTTTTT
651 TTTAGTTGCT CCTTTGATTC CTGAACATAA ACCTCAATTA CCTACACAAA
701 GTTGTGTGCT ATTCCGTTCC GGGGTAAATT CACAGTCTTC TAGTAGCTCT
751 TTATTCAAGT CCTACATGGT GCCTTATTTT TGGGAAGAAT TCGGGTTTCA
801 AAATAAGCAG CGTTTGTACA GTAATCACCA TATAGGGAGC CGTAATGGAT
851 TTTTACCTAC GTTGGTCTCT ATTCTTTGGG AACAAGACAA GGGGCCCTAT
901 CGTTCTCTATA TCTTTAAAGC AAAAGATTCT CAGGGCAATC CCCATCGCAT
951 AGGATTTTTA AGAATTTCTT CTTATGTTTG GACTGATTTA GAAGGACTTG
1001 AAGAGGATCA TAAGGATAGT CCTTGGGAGC TCTTTGGAGA GATCATCGAT

```

**Example 93**

The following *C.pneumoniae* protein (PID 4376424) was expressed <SEQ ID 185; cp6424>:

```

1 MMHNIVLSE EPGRSAFLGR TAFFPNKYPI AQGGVGIPST IGNLFTIWYC
51 FFYFRAATPQ SDHPDGCIFI LLERLKEGGA GFFYCDLRES NTTGFTLFFE
101 GSNKGVLEKH LFIRDE*

```

The cp6424 nucleotide sequence <SEQ ID 186> is:

```

1 ATGATGCACA ATATTGTTGT TCTTAGTGAG GAACCTGGAC GAAGCGCTTT
51 TCTTGGTAGG ACGGCATTTT TCCCTAATAA GTATCCAATA GCTCAGGGTG
101 GTGTTGGAAT ACCATCTACA ATAGGCAATC TCTTTACTAT ATGGTACTGT
151 TTCTATTTT ATAGAGCTGC AACTCCACAA TCTGATCATC CTGACGGATG
201 TGGCTTTATT CTACTAGAAA GGCTTAAGGA GCTCGGTGCA GGGTTCTTTT
251 ATTGTGATCT TCGTGAGTCC AATACCACTG GCTTTACTCT TTTTGTGAA
301 GGCTCCAATA AAGGTGTGTT AAAGAATCAC TTGTTTATTA GAGATGAGTA
351 A

```

The PSORT algorithm predicts cytoplasm (0.2502).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 93A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figure 93B) and for FACS analyses (Figure 93C; GST-fusion).

These experiments show that cp6424 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

**Example 94**

The following *C.pneumoniae* protein (PID 4376449) was expressed <SEQ ID 187; cp6449>:

```

1 VASETYPSTQI LHAQREVRDA YFNQADCHPA RANQILEAKK ICLLDVYHTN
51 HYSVFTFCVD NYPNLRFTFV SSKNNEMNGL SNPLDNVLE AMVRRTHARN
101 LLAACKIRNI EVPRVGLDL RSGILISKLE LKQPQFQSLT EDFVNHSTNQ
151 BEARVHQKHV LLISLILLCK QAVLESFQEK KRSS*

```

The cp6449 nucleotide sequence <SEQ ID 188> is:

```

1 GTGGCGTCTG AAACGTATCC TTCTCAGATA TTGCACGCTC AGAGGGAAGT
51 ACGTGATGCC TATTTTAATC AAGCGGATTG CCATCCTGCT CGGGCTAATC
101 AGATTCTCGA GGCTAAGAAA ATCTGTTTAT TAGATGTTTA TCATACTAAT
151 CATTATTCCG TATTTACTTT TTGTGTAGAT AATTATCCGA ATCTCCGCTT
201 TACATTTGTA TCTTCAAAAA ACAATGAGAT GAATGGCTTA TCTAATCCTC
251 TAGATAATGT TCTTGTAGAG GCTATGGTAC GTAGAACACA TGCAAGAAAC
301 CTACTTGACG CGTGTAATAAT TCGAAATATT GAGGTTCCAA GGGTTGTTGG
351 GCTTGACCTA AGATCTGGGA TACTCATTTT GAACTAGAA TTGAAGCAAC
401 CTCAGTTCCA AAGTTTAACA GAAGACTTCG TAAATCATTC CACAAATCAG
451 GAAGAAGCTC GCGTCCATCA AAAGCATGTG TTGCTAATTT CTTTAATTTT
501 ACTTTGCAAG CAGGCCGTTT TGAATCATTT CCAGGAAAAA AAGCGATCCT
551 CTTAA

```

The PSORT algorithm predicts inner membrane (0.2084).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 94A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figure 94B) and for FACS analyses (Figure 94C; GST-fusion).

These experiments show that cp6449 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.



```

1051 ATGGCAAGAA TCTTAGGCCA AGCCGAAGCC GAAGCTAATG AACTCTACAT
1101 GCGAGTCAAA CAACGTACCG ATGATGTAGA CACACTCACA GTCCTTATCT
1151 CTAAGATCAA TAATGAAAAG AAAGACATTG ATTGGAGTGA AAATGAAGAG
1201 ATGAAAGCTC TTTTAAATCG AGCTAAAGAG ATTGGAGTCA CTATAGACAA
1251 AGAAAAATAT ACTTGGACAG AAGAGGAAAA AAGACTTCTA AAAGAGAATG
1301 TCCAAATGCG CAAAGAGAAT ATGGAGAAAA TCACTCAAAT GGAAAGGACG
1351 GACATGCAAA GGCACCTCCA AGAGATTCTT CAATGTCATC AAGCGCGCTC
1401 TAATGTATTG AAGTTATTGA AAGAACTTAT GGACACCTTC ATTTACAACC
1451 TACGCCCTA A

```

10 The PSORT algorithm predicts cytoplasm (0.1308).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 91A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 91B) and for FACS analysis.

15 These experiments show that cp7353 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 92

The following *C.pneumoniae* protein (PID 4377408) was expressed <SEQ ID 183; cp7408>:

```

1 MLKIQKKRMC VSVVITVGAI VGFFNSADAA PKKKKIPIQI LYSFTKVSSY
51 LKNEDASTIF CVDVDRGLLQ HRYLGSPGWQ ETRRRQLFKS LENQSYGNER
101 LGEETLAIDI FRNKECLESE IPEQMEAILA NSSALVLGIS SFGITGIPAT
151 LHSLLRQNL S FQKRSIASES FLLKIDSAPS DASVFYKGV L FRGETAIVDA
201 LSQLFQAQLD SPKKIIFLGE DPEVVQAVGS ACIGWGMNFL GLVYYPQES
251 LFSYVHPYST ATELQEAQGL QVISDEVAQL TLNALPKMN*

```

The cp7408 nucleotide sequence <SEQ ID 184> is:

```

1 ATGTTGAAAA TCCAGAAAAA AAGAATGTGT GTCAGCGTAG TCATCACGGT
51 AGGCGCCATA GTGGGGTTTT TCAATTCTGC AGACGCAGCA CCAAAGAAAA
101 AGAAGATCCC TATACAGATT CTCTACTCCT TTAATAAAGT CTCTTCCTAT
151 TTAATAAACG AAGACGCAAG TACTATATTT TCGTTCGATG TGGATCGTGG
201 ACTTCTCCAG CATCGGTATT TAGGTAGTCC AGGATGGCAG GAAACCAGAC
30 251 GTCGGCAGTT ATTTAAATCC TTAGAAAATC AATCATACGG CAACGAACGT
301 TTAGGAGAAG AAACCTTTGC TATTGATATT TTCAGGAACA AAGAGTGCTT
351 GGAGAGCGAG ATCCCAGAGC AGATGGAAGC TATCCTTGCA AATTCTCGG
401 CCTTGGTCTT AGGCATCTCT TCTTTTGGGA TCACAGGAAT TCCTGCGACT
451 TTGCATAGTT TGCTTCGACA GAATCTATCT TTCCAAAAAC GCTCTATAGC
35 501 ATCGGAGAGC TTCTTTTAA AGATCGATAG TGCCCCCTCA GATGCCTCTG
551 TTTTATATAA AGGCGTGCTT TTCCGCGGAG AGACTGCGAT CGTGGATGCG
601 TTAAGCCAAT TATTTGCCCA GCTCGATCTT TCTCCTAAAA AAATTATCTT
651 TCTAGGAGAA GACCCGTAGG TCGTTCAAGC TGTGGGTCTT GCTTGATAG
701 GTTGGGGCAT GAACCTTTTA GGCCTGGTAT ACTATCCTGC TCAAGAAAGC
40 751 CTTTTTCTT ATGTTTCATC TTACTCTACA GCAACGGAGC TCCAAGAAGC
801 ACAGGGTTTA CAAGTAATTT CAGATGAAGT CGCACAGCTT ACTTTAAACG
851 CTCTTCCGAA AATGAATTAA

```

The PSORT algorithm predicts inner membrane (0.123).

45 The protein was expressed in *E.coli* and purified as a his-tag product (Figure 92A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 92B) and for FACS analysis.

These experiments show that cp7408 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

These experiments show that cp6506 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 97

The following *C.pneumoniae* protein (PID 4376882) was expressed <SEQ ID 193; cp6882>:

```

5      1  MSLNLNPSSQ DSASEDSTSQ SQIFDPIRNR ELVSTPEEKV RQRLLSFLMH
      51  KLNYPKKLII IEKELKTLFP LLMRKGLTLP KRRPDILIIIT PPTYTDAQGN
     101  THNLGDPKPL LLIECKALAV NQNALKQLLS YNYSIGATCI AMAGKHSQVS
     151  ALFNPKTQTL DFYPGLPEYS QLLNYFISLN L*

```

The cp6882 nucleotide sequence <SEQ ID 194> is:

```

10      1  ATGTCTTAT TGAACCTTCC CTCAAGCCAG GATTCTGCAT CTGAGGACTC
      51  CACATCGCAA TCTCAAATCT TCGATCCCAT TAGAAATCGG GAGTTAGTTT
     101  CTAATCCCGA AGAAAAAGTC CGCCAAAGGT TGCTCTCCTT CCTAATGCAT
     151  AAGCTGAACT ACCCTAAGAA ACTCATCATC ATAGAAAAAG AACTCAAAC
     201  TCTTTTTCCT CTGCTTATGC GTAAAGGAAC CCTAATCCCA AAACGCCGCC
     15  251  CAGATATTCT CATCATCACT CCCCCACAT ACACAGACGC ACAGGGAAAC
      301  ACTCACAACC TAGGCGACCC AAAACCCCTG CTACTTATCG AATGTAAGGC
     351  CTTAGCCGTA AACCAAAATG CACTCAAACA ACTCCTTAGC TATAACTACT
     401  CTATCGGAGC CACCTGCATT GCTATGGCAG GGAAACACTC TCAAGTGTC
     451  GCTCTCTTCA ATCCAAAAAC ACAAACTCTT GATTTTATAT CTGGCCTCCC
     20  501  AGAGTATTCC CAACTCCTAA ACTACTTTAT TTCTTTAAAC TTATAG

```

The PSORT algorithm predicts cytoplasm (0.362).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 97A). The protein was used to immunise mice, whose sera were used in a Western blot (Figure 97B) and for FACS analysis (Figure 97C).

25 These experiments show that cp6882 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 98

The following *C.pneumoniae* protein (PID 4376979) was expressed <SEQ ID 195; cp6979>:

```

30      1  MSVNPSGNSK NDLWITGAHD QHPDVKESGV TSANLGSHRV TASGGRQGLL
      51  ARIKEAVTGF FSRMSFFRSQ APRGSQQPSA PSADTVRSPL PGGDARATEG
     101  AGRNLIKKG YQGMKVITIP VPGGGAQRSS GSTTLKPTRP APPPPKTGGT
     151  NAKRPATHGK GPAPQPPKTG GTNAKRAATH GKGPAPQPPK GILKQPGQSG
     201  TSGKKRVSW S DED*

```

The cp6979 nucleotide sequence <SEQ ID 196> is:

```

35      1  ATGTCTGTTA ATCCATCAGG AAATTCCAAG AACGATCTCT GGATTACGGG
      51  AGCTCATGAT CAGCATCCCG ATGTTAAAGA ATCCGGGGTT ACAAGTGCTA
     101  ACCTAGGAAG TCATAGAGTG ACTGCCTCAG GAGGACGCCA AGGGTTATTA
     151  GCACGAATCA AAGAAGCAGT AACCGGGTTT TTTAGTCGGA TGAGCTTCTT
     201  CAGATCGGGA GCTCCAAGAG GTAGCCAACA ACCCTCTGCT CCATCTGCAG
     40  251  ATACTGTACG TAGCCCCTTG CCGGGAGGGG ATGCTCGCGC TACCGAGGGA
      301  GCTGGTAGGA ACTTAATTAA AAAAGGGTAC CAACCAGGGA TGAAAGTCAC
     351  TATCCACAG GTTCCTGGAG GAGGGGCCCA ACGTTCATCA GGTAGCACGA
     401  CACTAAAGCC TACGCGTCCG GCACCCCCAC CTCCTAAAC GGGTGGAAC
     451  AATGCAAAAC GTCCGGCAAC GCACGGGAAG GGTCCAGCAC CCCAGCCTCC
     45  501  TAAAACAGGT GGGACCAATG CTAAGCGCGC AGCAACGCAT GGGAAAGGTC
     551  CAGCACCTCA ACCTCCTAAG GGCATTTTGA AACAGCCTGG GCAGTCTGGG
     601  ACTTCAGGAA AGAAGCGTGT CAGCTGGTCT GACGAAGATT AA

```

The PSORT algorithm predicts cytoplasm (0.360).

**Example 95**

The following *C.pneumoniae* protein (PID 4376495) was expressed <SEQ ID 189; cp6495>:

MRELNAFELTQPEEYRNRWVLMPCCLKRCFRTQHAKVWSYRCVHEASLYEKNCFLLTYDDKHLPPQYGSILVKLHLQLFLKR  
LRKMISPHKIRYFECGAYGTLQRPYHLLLS

5 The cp6495 nucleotide sequence <SEQ ID 190> is:

TTGCGAGAATTAAATGCTTTTGAATTAACCTCAACCTGAAGAGTATCGAAACCGTTGGGTTTTGATGCCTTGTCTTAAGTGT  
CGTTTTTGTAGAACGCAACATGCAAAAGTCTGGTCTTATCGTTGTGTCCATGAAGCTTCTTTGTATGAGAAAAATTGTTTT  
CTTACTTTTGACTTATGATGATAAGCATTTACCTCAGTATGGTTCGTTGGTAAAGCTGCATTTACAGCTGTTTCTTAAGAGA  
TTAAGAAAGATGATTTCTCCTCATAAAATTCGTTATTTTGAATGTGGTGCATGGAACCAAATTACAAAGACCTCATTAT  
CATCTACTTTTATCATGA

10 The PSORT algorithm predicts cytoplasmic (0.280).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 95A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 95B) and for FACS analysis (Figure 95C).

15 These experiments show that cp6495 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

**Example 96**

The following *C.pneumoniae* protein (PID 4376506) was expressed <SEQ ID 191; cp6506>:

1 MRRFLFLILS SLPLVAFSAD NFTILEEKQS PLSRVSIIFA LPGVTFVSFD  
51 GNCPIPWFSH SKKTLEGQRI YYSGDSFGKY FVVSALWPNK VSSAVVACNM  
101 ILKHRVDLIL IIGSCYSRSQ DSRFGSVLVS KGYINYDADV RPFFERFEIP  
151 DIKKS VFATS EVHREAILRG GEEFISTHKQ EIEELLKTHG YLKSTTKTEH  
201 TLMGGLVATG ESFAMSRNYF LSLQKLYPEI HGFDSVSGAV SQVCYEYSIP  
251 CLGVNILLPH PLESRSNEDW KHLQSEASKI YMDTLLKSVL KELCSSH\*

25 The cp6506 nucleotide sequence <SEQ ID 192> is:

1 ATGCGTCGTT TTCTGTTTCT TATTCTTAGC TCTCTTCCTT TGGTCGCATT  
51 CTCTGCTGAT AATTTCACTA TTCTAGAAGA AAAACAGAGT CCTTTAAGTC  
101 GTGTAAGTAT TATTTTGTCT TTACCTGGGG TTACTCCCGT TTCTTTTGAT  
151 GGTAAATGTC CTATTCTCTG GTTTTCTCAT AGTAAAAAGA CTCTAGAGGG  
30 201 ACAGAGAATT TATTACTCTG GCGACTCCTT TGGGAAATAC TTTGTAGTTT  
251 CTGCTCTTTG GCCTAATAAA GTTTCTTCAG CTGTGTGGC TTGTAATATG  
301 ATTCTTAAAC ATCGAGTGGA TCTTATTCTA ATTATAGGCT CGTGTACTCT  
351 TAGGTCTCAA GATAGCCGTT TTGGCAGCGT CTTAGTTTCT AAAGGCTACA  
401 TTAATTATGA TGCAGATGTG AGGCCTTTCT TTGAAAGATT TGAGATTCCA  
35 451 GACATTAAAA AGAGTGTTTT TGCAACCAGT GAGGTTTCATC GGGAGGCAAT  
501 TCTTCGTGGA GGCGAAGAGT TTATTCTTAC CCATAAACAA GAAATCGAAG  
551 AGCTTTTGAA GACTCATGGG TATTTGAAAT CAACAACCAA AACGGAGCAC  
601 ACCTTAATGG AAGGTTTGGT TGCTACAGGC GAGTCTTTCG CGATGTCGGC  
651 AAACCTATTTT CTTTCTCTAC AAAAATGTGA TCCAGAGATT CATGGTTTGT  
40 701 ATAGTGTCAG CGGCGCTGTT TCTCAGGTAT GCTATGAATA TAGCATTCCT  
751 TGTTTAGGTG TGAATATCCT TCTCCCTCAT CCTTTAGAAT CACGGAGTAA  
801 CGAGGATTGG AAGCATCTTC AAAGTGAGGC AAGTAAAATT TATATGATA  
851 CCTTGCTCAA GAGTGTATTA AAAGAACTCT GTTCTTCTCA TTAA

The PSORT algorithm predicts periplasmic space (0.571).

45 The protein was expressed in *E.coli* and purified as his-tag (Figure 96A) and GST-fusion (Figure 96B) products. The GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 96C) and for FACS analysis (Figure 96D).

251 AATGTTATAC CCGATTTGAA GATGGCACAA TTTTTTATGA ATGCGATTAG

The PSORT algorithm predicts inner membrane (0.143).

The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 100A) and a his-tag product. The proteins were used to immunise mice, whose sera were used in a Western blot (Figure 100B) and for FACS analysis (Figure 100C).

These experiments show that cp7355 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 101

The following *C.pneumoniae* protein (PID 4377380) was expressed <SEQ ID 201; cp7380>:

```

10      1  VHYCERTLDP KYILKIALKL RQSLSLFFQN SQSLQRAYST PYSYRRIILQ
      51  KENKEKQALA RHKCISILEF FKNLLFVHLL SLSKNQREGC STDMAVVSTP
     101  FFNRLNLWYRL LSSRFSLWKS YCPRFFLDYL EAFGLLSDFL DHQAVIKFFE
     151  LETHFSYYPV SGFVAPHQYL SLLQDRYFPI ASVMRTLDKD NFSLTPDLIH
     201  DLLGHVPWLL HPSFSEFFIN MGRFLTQVIE KVQALPSKKQ RIQTLQSNLI
     15  251  AIVRCFWFTV ESGLIENHEG RKAYGAVLIS SPQELGHAFI DNVRLPLEL
     301  DQIIRLPFNT STPQETLFSI RHFDELVELT SKLEWMLDQG LLESIPLYNQ
     351  EKYLSGFEVL CQ*
```

The cp7380 nucleotide sequence <SEQ ID 202> is:

```

20      1  GTGCACTACT GCGAGAGAAC CCTGGACCCA AAGTATATTC TGAAGATTGC
      51  TCTAAAGCTG AGACAATCAC TTTCCCTGTT CTTCCAGAAC AGCCAATCAC
     101  TCCAACGTGC ATACTCGACC CCATATTCCCT ACTACCGAAT CATTCTACAA
     151  AAGGAAAATA AAGAGAAGCA AGCTTTAGCT CGACACAAAT GCATTTCTAT
     201  TTTAGAATTT TTCAAAACT TACTCTTTGT TCATCTTCTG TCATTATCAA
     25  251  AGAATCAAAG GGAAGGTTGC TCCACTGATA TGGCTGTTGT AAGCACTCCC
     301  TTTTTTAATC GGAATTTATG GTATCGACTC CTTTCCTCAC GGTTTTCTCT
     351  ATGGAAGAGC TATTGTCCAA GATTTTCTCT TGATTACTTA GAAGCTTTTCG
     401  GTCTCCTTTC TGATTTCTTA GACCATCAAG CAGTCATTAA ATTCTTCGAA
     451  TTAGAAACAC ATTTTTCCTA TTATCCCGTT TCAGGATTG TAGCTCCCCA
     501  TCAATACTTG TCTCTGTTGC AGGACCGTTA CTTTCCCATT GCCTCTGTAA
     30  551  TGCGAACTCT CGATAAAGAT AATTTCTCCT TAACTCCTGA TCTCATCCAT
     601  GACCTTTTAG GGCACGTGCC TTGGCTTCTA CATCCCTCAT TTTCTGAATT
     651  TTTCAATAAC ATGGGAAGAC TCTTCACTAA AGTCATAGAA AAAGTACAAG
     701  CTCTTCCTAG TAAAAACAA CGCATACAAA CCCTACAAAG CAATCTGATC
     751  GCTATTGTAC GCTGCTTTTG GTTTACTGTT GAAAGCGGAC TTATTGAAAA
     35  801  CCATGAAGGA AGAAAAGCAT ATGGAGCCGT TCTTATCAGT TCTCCTCAGG
     851  AACTTGGACA CGCTTTCATT GATAACGTAC GTGTTCTCCC TTTAGAATTG
     901  GATCAGATTA TTCGTCTTCC CTTCAATACA TCAACTCCAC AAGAGACTTT
     951  ATTTTCAATA AGACATTTTG ATGAAGTGGT AGAACTCACT TCAAAATTAG
    1001  AATGGATGCT CGACCAAGGT CTGTTAGAAT CAATTCCCCT TTACAATCAA
     40  1051  GAGAAATATC TTTCTGTTT TGAGGTACTT TGCCAATGA
```

The PSORT algorithm predicts inner membrane (0.1362).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 101A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 101B) and for FACS analysis (Figure 101C).

45 These experiments show that cp7380 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 102

The following *C.pneumoniae* protein (PID 4376904) was expressed <SEQ ID 203; cp6904>:

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 98A). The GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 98B) and for FACS analysis (Figure 98C).

These experiments show that cp6979 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 99

The following *C.pneumoniae* protein (PID 4377028) was expressed <SEQ ID 197; cp7028>:

```

1  MLLGFLCDP CASWQCAAVA NCYDSVFMSR PEHKPNIPYI TKATRRGLRM
51  KTLAYLASLK DARQLAYDFL KDPGSLARLA KALIAPKEAL QEGNLFYFGC
101 SNIEDILEEM RRPHRILLLG FSYCQKPKAC PEGRFNDACR YDP SHPTCAS
151 CSIGTMMRLN ARRYTTVIIP TFIDIAKHLH TLKKRYPGYQ ILFAVTACEL
201 SLKMFQDYAS VMNLKGVGIR LTGRICNTFK AFKLAERGVK PGVTILEEDG
251 FEVLARILTE YSSAPFPRDF CEIH*

```

The cp7028 nucleotide sequence <SEQ ID 198> is:

```

15 1  ATGCTTCTAG GGTTTTTGTG TGA CTGCCCC TGTGCTTCGT GGCAGTGTGC
51 51  GGCCGTGCT AATTGTTATG ATTCCGTATT TATGTCTAGA CCAGAGCACA
101 101 AACCTAATAT TCCTTATATT ACTAAAGCTA CAAGACGGGG TCTGCGTATG
151 151 AAGACGCTTG CTTATCTGGC CTCTTTAAAA GATGCTAGAC AGCTTGCCTA
201 201 TGATTTCTG AAAGATCCTG GTTCTTTAGC TCGGTTAGCT AAGGCTTTGA
20 251 TAGCTCCTAA GGAGGCCTTA CAGGAGGGCA ACCTATTTTT TTATGGCTGT
301 301 AGTAATATTG AGGATATTTT AGAGGAGATG CGTCGTCCTC ATAGAATCCT
351 351 TTTGTTAGGA TTTTCTTATT GTCAAAAGCC TAAGGCATGT CCTGAAGGGC
401 401 GTTTC AATGA TGCTTGTGGG TATGATCCTT CACATCCTAC ATGTGCCTCA
451 451 TGTCTATAG GGACCATGAT GCGGCTGAAT GCTCGTAGAT ACACTACTGT
25 501 501 GATCATCCCT ACATTTATAG ATATCGCAA ACATTTACAC ACTTTAAAAA
551 551 AGCGCTACCC TGGATATCAA ATTCTCTTTG CAGTTACTGC TTGTGAACCTT
601 601 TCCTTAAAAA TGTTTGGAGA TTATGCCTCC GTAATGAACT TAAAGGGTGT
651 651 GGGCATCAGA CTCACAGGAC GTATTTGCAA TACATTTAAG GCATTTAAAT
701 701 TAGCTGAGCG AGGAGTCAA CCAGGAGTCA CTATCCTAGA AGAAGATGGC
30 751 751 TTGAGGTAT TAGCAAGGAT TCTTACAGAA TACAGTAGCG CTCCTTTCCC
801 801 TAGAGACTTT TGTGAGATCC ATTAG

```

The PSORT algorithm predicts cytoplasm (0.1453).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 99A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 99B) and for FACS analysis (Figure 99C).

These experiments show that cp7028 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 100

The following *C.pneumoniae* protein (PID 4377355) was expressed <SEQ ID 199; cp7355>:

```

40 1  MKKVVTLSII FFATYCASEL SAVTVVAVPL SEAPGKIQVR PVVGLQFQEE
51 51  QGSVPYSFY PYDYGYYYPE TYGYTKNTGQ ESRECYTRFE DGTIFYECD*

```

The cp7355 nucleotide sequence <SEQ ID 200> is:

```

45 1  ATGAAGAAAG TCGTAACACT ATCCATTATA TTTTTCGCAA CGTATTGTGC
51 51  ATCAGAGCTT AGTGCTGTAA CTGTAGTGGC TGTGCCTTTA TCAGAGGCTC
101 101 CAGGGAAGAT TCAAGTTCGT CCCGTCGTTG GTCTGCAATT TCAAGAAGAA
151 151 CAGGTTCTG TGCCCTATAG TTTTATTAT CTTATGACT ATGGGTATTA
201 201 CTATCCAGAG ACTTATGGCT ATACTAAAAA TACAGGTCAA GAAAGTCGCG

```

-136-

```

1  LNFAKIDHNH  LYLTCIGDLG  VACPILSTDC  LPNYSEKASH  EVLVYSKFRC
51  ISGEP SRLAT  SGNDTYYSIV  SLPIGLRYEV  TSPSGRHDFN  IDMHVAPKIG
101 AVLSHGTREA  KEIPGSSKDY  AFFSLTARES  LMISEKLAMT  FQVSEVIQNC
151 YSQCTKVTKT  NLKEQYRHLS  HNTGFELSVK  SAF*

```

5 The cp7387 nucleotide sequence <SEQ ID 208> is:

```

1  TTGAATTTTG  CAAAGATTGA  TCACAATCAT  CTCTACCTTA  CATGTTTGGG
51  AGATCTTGGT  GTAGCTTGTC  CTATACTTTC  TACAGATTGT  CTACCTAATT
101 ATAGCGAGAA  AGCATCTCAT  GAGGTCTTTG  TTTATAGTAA  ATTTAGATGC
10  151 ATTTCTGGAG  AGCCATCTCG  ACTTGCAACT  TCAGGAAATG  ACACATATTA
201 TTCTATAGTA  AGTTTACCTA  TAGGACTCCG  TTACGAAGTG  ACTTCACCAT
251 CAGGACGTCA  TGATTTC AAT  ATTGATATGC  ATGTAGCTCC  AAAGATAGGT
301 GCAGTACTCT  CTCATGGAAC  ACGAGAGGCT  AAAGAGATCC  CAGGATCTTC
351 AAAAGACTAT  GCATTTT TTA  GCTTGACTGC  TAGAGAAAGT  TTAATGATTT
401 CTGAAAAGCT  TGCGATGACT  TTCCAAGTTA  GCGAAGTTAT  TCAGAATTGT
15  451 TATTCACAAT  GTACTAAAGT  AACGAAAAGT  AATTAAAAAG  AACAGTATAG
501 GCACTTATCC  CACAATACAG  GGTTTGAGTT  AAGCGTCAAG  TCTGCATTCT
551 AA

```

The PSORT algorithm predicts inner membrane (0.043).

The protein was expressed in *E.coli* and purified as a his-tagged-fusion product (Figure 104A) and also as a GST-fusion (Figure 104B). The recombinant proteins were used to immunise mice, whose sera were used in a Western blot and for FACS analysis (Figure 104C; his-tagged).

These experiments show that cp7387 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 105

25 The following *C.pneumoniae* protein (PID 4376281) was expressed <SEQ ID 209; cp6281>:

```

1  MFLQFFHPIV  FSDQSLSFLP  YLGKSSGIIE  KCSNIVEHYL  HLGGDTSVII
51  TGVSGATFLS  VDHALPISKS  EKIIKILSYI  LILPLILALF  IKIVLRILF
101 FKYRGLILDV  KKEDLKKTLT  PDQENLSLPL  PSPPTLKKIH  ALHILVRSGK
30  151 TYNELIQEGF  SFTKITDLGQ  APSPKQDIGF  SYNSSLNPNFY  FHSLSVSPNI
201 SGEERALNYH  KEQQEEMAVK  LKTMQACSFV  FRSLHLPSMQ  TKDKKAGFGL
251 LTFPPWKIYP  L*

```

The cp6281 nucleotide sequence <SEQ ID 210> is:

```

1  ATGTTTCTTC  AGTTT TTTCA  TCCTATAGTC  TTC TCGGATC  AGTCCTTATC
51  TTTTCTTCCT  TACCTAGGAA  AAAGCTCTGG  CATTATTGAA  AAATGTTCCA
35  101 ATATCGTTGA  ACACTATTTA  CATTTGGGAG  GAGACACTTC  TGTATCATC
151 ACAGGAGTTT  CTGGAGCTAC  CTTTCTATCT  GTTGATCATG  CCCTCCCAAT
201 CTCGAAATCT  GAAAAAATAA  TAAAAATTCT  CTCCTATATT  TTAATTCTTC
251 CTCTGATTCT  AGCTCTCTTT  ATTAAGATCG  TTTTACGCAT  TATCTTATTC
301 TTCAAGTATC  GTGGTCTAAT  CCTAGATGTT  AAGAAGGAGG  ATTTGAAAAA
40  351 AACACTTACA  CCTGACCAAG  AAAACCTCAG  TCTTCCTTTA  CCATCTCCTA
401 CAACATTAAA  GAAAATTCAT  GCGCTACACA  TTTTAGTGCG  TTCTGGAAAA
451 ACCTATAACG  AGCTTATACA  AGAAGGGTTT  TCTTTCAC TA  AAATCACAGA
501 TCTTGGTCAA  GCTCCTTCAC  CAAAGCAAGA  TATTGGCTTC  TCTTATAATT
551 CCCTTCTCCC  TAACTTCTAT  TTTCATTCCCT  TGGTATCTGT  TCCAAATATT
45  601 TCAGGCGAGG  AACGGGCTCT  TAATTATCAT  AAAGAACAAC  AAGAGGAAAT
651 GGCTGTTAAA  TTAAAACAA  TGCAAGCGTG  TTCTTTTGTC  TTCCGATCCC
701 TGCATTTACC  TTCAATGCAA  ACGAAGGACA  AAAAGGCTGG  ATTTGGACTA
751 CTGACGTTT  TCCCTTGGA  AATCTACCCC  CTATAA

```

The PSORT algorithm predicts inner membrane (0.5373).

50 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 105A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 105B) and for FACS analysis.

1 MMNYEDAKLR GQAVAILYQI GAIKFGKHIL ASGEETPLYV DMRLVISSPE  
 51 VLQTVATLIW RLRPSFNSSL LCGVPYTALT LATSISLKYN IPMVLRRKEL  
 101 QNVDPSSDAIK VEGLFTPGQT CLVINDMVSS GKSIIETAVA LEENGLVVRE  
 151 ALVFLDRRKE ACQPLGPQGI KVSSVFTVPT LIKALIAYGK LSSGDLTLAN  
 201 KISEILEIES \*

The cp6904 nucleotide sequence <SEQ ID 204> is:

1 ATGATGAACT ACGAAGATGC AAAATTACGC GGTCAAGCTG TAGCAATTCT  
 51 ATACCAAATC GGAGCTATAA AGTTCGGAAA ACATATTCTC GCTAGCGGAG  
 101 AAGAAACTCC TCTGTATGTA GATATGCGTC TTGTGATCTC CTCTCCAGAA  
 151 GTTCTCCAGA CAGTGGCAAC TCTTATTTGG CGCTCCGCC CCTCATTCAA  
 201 TAGTAGCTTA CTCTGCGGAG TCCCTTATAC TGCTCTAACC CTAGCAACCT  
 251 CGATCTCTTT AAAATATAAC ATCCCTATGG TATTGCGAAG GAAGGAATTA  
 301 CAGAATGTAG ACCCCTCGGA CGCTATTAAA GTAGAAGGGT TATTTACTCC  
 351 AGGACAAACT TGTTTAGTCA TCAATGATAT GGTTCCTCA GGAATCTA  
 401 TAATAGAGAC AGCAGTCGCA CTGGAAGAAA ATGGTCTGGT AGTTCGTGAA  
 451 GCATTGGTAT TCTTAGATCG TAGAAAAGAA GCGTGTCAAC CACTTGGTCC  
 501 ACAGGGAATA AAAGTCAGTT CGGTATTTAC TGTACCCACT CTGATAAAG  
 551 CTTTGATCGC TTATGGGAAG CTAAGCAGTG GTGATCTAAC CCTGGCAAAC  
 601 AAAATTTCCG AAATTCTAGA AATTGAATCT TAA

The PSORT algorithm predicts cytoplasm (0.0358).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 102A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 102B) and for FACS analysis.

The cp6904 protein was also identified in the 2D-PAGE experiment.

These experiments show that cp6904 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 103

The following *C.pneumoniae* protein (PID 4376964) was expressed <SEQ ID 205; cp6964>:

1 MKKLIALIGI FLVPIKNTN KEHDAHATVL KAARAKYNLF FVQDVPVHE  
 51 VIEPISPDCL VHYEGWV\*

The cp6964 nucleotide sequence <SEQ ID 206> is:

1 ATGAAAAAAT TGATTGCTTT GATAGGGATA TTTCTTGTTT CAATAAAAGG  
 51 AAATACCAAT AAGGAACACG ACGCTCACGC GACTGTTTTA AAAGCGGCCA  
 101 GAGCAAAGTA TAATTTGTTT TTTGTTTCAAG ATGTTTTCCC TGACACGAA  
 151 GTTATCGAGC CTATTTCTCC CGATTGCCTG GTACATTATG AAGGGTGGGT  
 201 TTGA

The PSORT algorithm predicts inner membrane (0.091).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 103A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 103B) and for FACS analysis (Figure 103C).

These experiments show that cp6964 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 104

The following *C.pneumoniae* protein (PID 4377387) was expressed <SEQ ID 207; cp7387>:

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 108A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 108B) and for FACS analysis.

These experiments show that cp7400 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 109

The following *C.pneumoniae* protein (PID 4376395) was expressed <SEQ ID 217; cp6395>:

```

1  MENAMSSSFV YNGPSWILKT SVAQEVFKKH GKGIQVLLST SVMFLFIGLV
51  CAFIFPQYLI VFLVTIALLM LAISLVLFLL IRSVRSSMVD RLWCSEKGYA
101 LHQHENGPFLL DVKRVQQILL RSPYIKVRAL WPSGDIPEDP SQAAVLLLSL
151 WFFFSSVDVE ALLPSPQKEKE GKYIDPVLPLK LSRIERVSLV VFLSAFTLDD
201 LNEQGVNPLM NNEEFLLFFIN KKAREHGIQD LKHEIMSSLE KTGVPPLDPSM
251 SFQVSQAMFS VYRYLRQRDL TTSELRCFHL LSCFKGDVVH CLASFENPKD
301 LADSDFLEAC KNVEWGEFIS ACEKALLKNP QGISIKDLKQ FLVR*

```

The cp6395 nucleotide sequence <SEQ ID 218> is:

```

1  ATGGAGAATG CTATGTCATC ATCGTTTGTG TATAATGGGC CTTCTGGGAT
51  TTTAAAAACG TCAGTAGCTC AGGAGGTATT TAAAAAGCAC GGTAAGGGGA
101 TTCAGGTTCT CTTAAGTACT TCAGTGATGC TTTTATAGG TCTTGGAGTC
151 TGTGCCTTTA TATTTCTCTA ATATCTGATT GTTTTGTGTT TGAATATAGC
201 TTTGCTTATG CTCGCTATAA GCTTGGTATT GTTCTCTCTA ATACGTTCTG
251 TACGCTCTTC AATGGTAGAT CGTTTGTGGT GTTCTGAAAA AGGATATGCT
301 CTTTCATCAAC ATGAGAACGG GCCTTTTGTG GATGTGAAGC GTGTACAGCA
351 AATTCTTCTA AGATCACCCCT ATATTAAAGT TCGGGCTTTA TGGCCGCTCTG
401 GAGATATCCC TGAGGATCCT TCACAAGCTG CGGTTCTATT ACTTTCTCCT
451 TGGACTTTCT TTTTCATCCGT GGATGTAGAG GCTTTATTAC CGAGTCCTCA
501 AGAAAAGGAG GGTAAGTATA TAGATCCTGT GCTGCCTAAG TTGTCTAGGA
551 TAGAGAGAGT CTCACCTTTA GTGTTTTTGA GTGCATTTAC TTTGGATGAC
601 TTAAACGAAC AGGGAGTCAA TCCTTTGATG AATAATGAGG AATTTTATT
651 TTTTATAAAT AAGAAAGCGC GTGAGCATGG GATTCAGGAT TTAAACACG
701 AGATTATGTC TTCGTTAGAG AAAACAGGAG TGCCATTAGA CCCCTCAATG
751 AGTTTTCAAG TTTCACAAGC GATGTTTTCT GTATATCGCT ACTTGAGACA
801 AAGGGATTTA ACGACTTCAG AATTAAGATG TTTTCACCTC TTAAGTTGTT
851 TTAAAGGGGA TGTGGTTCAT TGTTTAGCTT CATTTGAAAA CCCTAAAGAT
901 TTAGCAGATT CTGACTTTT AGAAGCTTGT AAGAACGTGG AATGGGGTGA
951 GTTTATTTTC GCATGTGAGA AGGCTCTTTT AAAGAATCCG CAAGGAATTT
1001 CCATTAAGGA TCTAAAACAA TTTTATAGTA GGTAA

```

The PSORT algorithm predicts inner membrane (0.6307).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 109A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 109B) and for FACS analysis.

These experiments show that cp6395 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 110

The following *C.pneumoniae* protein (PID 4376396) was expressed <SEQ ID 219; cp6396>:

```

1  MIEFAFVPHS SVTADRIEDR MACRMNKLST LAITSLCVLI SSVCMIGIL
51  CISGTVGTYA FVVGIIFSLV ALVACVFFLY FFYFSSEEFK CASSQEFRLF
101 PIPAVVSALR SYEYISQDAI NDVIKDTMQL STLSSLLDPE AFFLEFPYFN
151 SLIVNHSMEKE ADRLSREAFLL ILLGEITWKD CETKILPWLK DPNITPDDFW
201 KLLKDHFDLKL DFKKRIATWI RKAYPEIRLP KKHCLDKSTY KGCKKFLLLS

```



These experiments show that cp6281 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 106 and Example 107

- 5 The following *C.pneumoniae* protein (PID 4376306) was expressed <SEQ ID 211; cp6306>:

```

1  MGNHETYIHP GVLPSHQAQD VSRSTVYPSR SFIMRRMLMG WNFNRVPSKS
51  SEQLMDGHRI PLIFFGKHPH TISILNVNRF SWLSIFYNGE RGF*

```

The cp6306 nucleotide sequence <SEQ ID 212> is:

```

10      1  ATGGGAAACC ATGAGACCTA TATACATCCA GGAGTGCTCC CGAGTAGTCA
      51  TGCTCAGGAT GTTAGCAGAT CTACAGTTTA CCCCAGTCGA AGTTTTATCA
      101  TGAGACGTAT GCTCATGGGC TGAATTTC AATCGTGTTC CTCGAAGAGC
      151  TCCGAGCAGT TAATGGATGG TCATCGCATA CCTCTTATAT TTTTGGGAA
      201  GCATCATCCT ACTATATCTA TTTTAAATGT CAATAGATTT TCTTGGCTCT
      251  CCATTTTTTA CAATGGAGAA AGGGGGTTTT GA

```

- 15 The PSORT algorithm predicts cytoplasm (0.167).

The following *C.pneumoniae* protein (PID 4376434) was also expressed <SEQ ID 213; cp6434>:

```

1  MSIESINRSIH LEASTPFFIK LTNLCESRLV KITSLSVISLL ALVGAGVTLV
51  VLFVAGILPL LPVLILEIIL ITVLVLLFCL VLEPYLIEKP SKIKELPKVD
101  ELSVVETDST L*

```

- 20 The cp6434 nucleotide sequence <SEQ ID 214> is:

```

      1  ATGTCTGAAA GTATTAACAG AAGCATTCAT TTAGAAGCCT CTACACCATT
      51  TTTTATAAAA TTAACGAATC TCTGTGAAAG TAGATTAGTT AAGATCACTT
      101  CTCTTGTTAT TTCTCTATTA GCTTTAGTGG GTGCGGGAGT CACTCTTGTTG
      151  GTTTTATTTG TAGCTGGGAT CCTTCCTTTA CTTCTGTAC TCATCTTAGA
      201  AATTATTTTA ATAACCGTCC TTGTCTTGCT TTTTGTGTTG GTATTGGAAC
      251  CTTATTTAAT AGAAAAACCT AGTAAAATAA AGGAACTACC TAAAGTAGAC
      301  GAGCTATCTG TAGTAGAAAC GGACAGTACT CTTTAA

```

The PSORT algorithm predicts inner membrane (0.6859).

- 30 The proteins were expressed in *E.coli* and purified as his-tag products (Figure 106A; 6306 = lanes 2-4; 6434 = lanes 8-10). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 106B & 107) and for FACS analysis.

These experiments show that cp6306 & cp6434 are surface-exposed and immunoaccessible proteins, and that they are useful immunogens. These properties are not evident from the sequences alone.

#### Example 108

- 35 The following *C.pneumoniae* protein (PID 4377400) was expressed <SEQ ID 215; cp7400>:

```

1  MRVMRFFCLF FLGFLGSFHC VAEDKGVDLF GVWDDNQITE CDDSYMTEGR
51  EEVEKVVD

```

The cp7400 nucleotide sequence <SEQ ID 216> is:

```

40      1  GTGAGAGTTA TGAGATTTTT TTGTCTATTT TTTCTTGGGT TCCTAGGATC
      51  TTTTCATTGT GTTGCTGAAG ACAAGGGCGT GGATTATTTT GGAGTCTGGG
      101  ACGATAACCA AATTACAGAG TGTGACGATA GTTACATGAC AGAGGGTCGT
      151  GAAGAGGTTG AAAAGGTAGT GGACGCTTAG

```

The PSORT algorithm predicts periplasmic space (0.924).

-140-

```

751  AACTACCACT CAAAATCTTT TGCTAGTGGT AGTTATGACT TTATTGCAAA
801  GCCCCTATTC GAACAAACAA ATGTAGACGG CTACTATTTA GAGTTTGATC
851  ATGAGCGTTC TGGAGACTTC TCTCCTCTCA CCTTCATTTT TGGAGAAAAA
901  ACTGTCTGCT TAGGTCTTGT TACCAGCAAA ACCCCTACAC TTGAAAATAA
951  GGATGAGGTC ATTGCTCGCA TACATCAAGC AGCAGACTAC CTGCCCTTGG
1001 AAAGACTCTC TCTAAGTCCA CAGTGTGGTT TTGCTTCATG TGAAATAGGA
1051 AATAAATTAA CAGAAGAAGA GCAATGGGCT AAAGTTGCTC TAGTAAAAGA
1101 AATTTCCGAA GAAGTTTGGA AATAA

```

The PSORT algorithm predicts cytoplasm (0.2171).

- 10 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 111A) and also as a his-tagged product. The his-tag protein was used to immunise mice, whose sera were used in a Western blot (Figure 111B) and for FACS analysis.

These experiments show that cp6408 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## 15 Example 112

The following *C.pneumoniae* protein (PID 4376430) was expressed <SEQ ID 223; cp6430>:

```

1  MKLYSISSDV DTPWIFQLMS KVDSYLFLLG NRIKVVSIVM QEPNLIIGKV
51  ENVRISTIVK ILKILSFLIF PLILIALALH YFLHAKYANH LLVSKILERA
101 PQYVPIGRS GDTASHYKLT TLVPSQKNL QAMGSNPLEV EAALRTTKPS
151 FFCVPAKYRQ IISSHGIRF SLDLEQLADD INLDSVSWPT EYLNSTMDFC
201 SKADKRVIQN VQNLRTGTIY NSVGKRSLK FMLQHLFIDG ITQENPEALP
251 NNTSGRLTLF PSVRYIYSHF TPQNPTIWPQ VFFRQGPLDE DRGGGFELIE
301 QLQELGVRFP ICPSQGPDPN NFQGFQGIRI YWEDSYQPNK EV*

```

The cp6430 nucleotide sequence <SEQ ID 224> is:

```

25 1  ATGAAACTTT ATAGCATCTC TTCAGATGTA GATACACCTT GGATATTTCA
51  GCTTATGTCA AAGGTAGATT CTTATCTTTT CTTAGGCGGG AATAGAAATCA
101 AGGTTGTATC TATAGTTATG CAAGAACCTA ACTTAATTAT TGGAAAAGTA
151 GAAAACGTTT GGATCTCCAC AATAGTGAAA ATATTAAAGA TTTTATCCTT
201 CTTAATCTTC CCTCTGATTT TAATCGCTTT AGCCCTACAC TATTTTCTAC
30 251 ATGCTAAATA TGCTAATCAC TTACTTGTAT CTAAGATTTT AGAAAGAGCT
301 CCTCAGTATG TGCCTATTC TGGTCGTTCA GGAGACACGG CGTCTCATT
351 TAAATTAACA ACATTGGTTC CAGTATCCCA AAAAAATCTA CAAGCTATGG
401 GATCAAATCC TCTAGAAGTT GAAGCGGCTC TTCGAACCTA AAAACCTCT
451 TTTTCTGTG TACCTGCAAA ATACCGTCAG ATTATAATTT CAAGTCACGG
35 501 CATTCGCTTT TCTTTAGATC TTGAACAAC TGTGATGAC ATTAATTTAG
551 ATTCGGTTTC CTGGCCTACG GAGTATCTTA ACTCTACTAT GGATTTTTCG
601 AGCAAGGCAG ATAAACGTGT TATACAGAAT GTACAAAATC TCGGACACGG
651 AACTTACATA AATTCTGTAG GAAAGCGTAG CCTTTTAAAA TTCATGTTAC
701 AGCACCTATT TATTGATGGG ATCACACAAG AAAACCTGA AGCCCTTCCT
40 751 AACAATACAT CTGGAAGACT GACTCTATTC CCTAGTGTTC GTTATATCTA
801 TTCTCATTTT ACTCCACAAA ATCCTACAAT ATGGCCGCAA GTCTTTTTC
851 GACAAGGTCC TCTAGATGAA GATCGAGGAG GAGGATTGTA GATCTTAGAG
901 CAATTACAAG AGTTAGGAGT TAGGTTTCCA ATTTGCCCTT CTCAAGGACC
951 AGACAATCCT AATTTTCAAG GTTTTCAAGG GATTCGTATC TATTGGGAAG
45 1001 ATTCTATCA ACCCAATAAG GAGGTTTAA

```

The PSORT algorithm predicts inner membrane (0.5140).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 112A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 112B) and for FACS analysis.

- 50 These experiments show that cp6430 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

251 ENDVQYQRLH HKVCYFSGEF PAMVLGLGSE VPMVLGLPKV PKDLTWEMFM  
301 ENMPVLLQSK REGHWKISLE DVASL\*

The cp6396 nucleotide sequence <SEQ ID 220> is:

```

5      1  ATGATCGAGT TTGCTTTTGT TCCTCATACC TCCGTGACAG CGGATCGGAT
      51  TGAGGATCGC ATGGCCTGTC GCATGAACAA GTTGTCTACT TTAGCAATTA
     101  CAAGTCTTTG TGTATTGATC AGTTCAGTTT GTATTATGAT TGGGATTTTA
     151  TGCATTTCTG GAACGGTTGG GACCTATGCA TTTGTTGTAG GAATTATTTT
     201  TTCTGTGCTT GCTTTGGTAG CATGTGTTTT CTTCTTTTAT TTCTTTTATT
     251  TTTCTTCTGA GGAATTTAAG TGTGCTTCTT CGCAGGAGTT TCGTTTTTTG
     301  CCTATACCAG CTGTGGTTTC TGCATTGCGT TCCTATGAAT ACATTTCTCA
     351  GGACGCTATC AATGACGTTA TAAAAGATAC GATGCAGTTG TCTACCCTTT
     401  CTTCTCTTTT AGATCCCGAA GCTTTTTTCT TAGAATTTCC TTATTTTAAC
     451  TCTTTGATAG TGAATCATTC GATGAAGGAA CGCGATCGTT TGTCTCGAGA
     501  GGCTTTTTTG ATTTTATTAG GTGAGATTAC TTGGAAGGAT TGTGAAACAA
     551  AAATTTTGCC ATGGTTGAAA GATCCTAATA TCACTCCTGA TGATTTCTGG
     601  AAGCTATTAA AAGACCATTT CGATTTAAAG GACTTTAAGA AGAGGATCGC
     651  CACTTGGAATA CGGAAGGCCT ATCCAGAAAT TAGATTACCG AAGAAGCATT
     701  GTTTAGATAA GTCTATCTAT AAGGGGTGTT GTAAGTTTTT ATTACTTTCT
     751  GAGAATGATG TGCAATATCA GAGGTATTAT CATAAGGTCT GTTATTTCTC
     801  TGGGGAGTTT CCTGCCATGG TTTTAGGTTT GGGAAAGTGA GTGCCTATGG
     851  TGTTAGGACT CCCTAAGGTT CCCAAGGATC TTACCTGGGA GATGTTTATG
     901  GAAAATATGC CTGTTCTTCT GCAAAGCAAA AGAGAGGGGC ATTGGAAAAA
     951  CTCCTTGGA GACGTAGCCT CTCTTTAA

```

The PSORT algorithm predicts inner membrane (0.6095).

- 25 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 110A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 110B) and for FACS analysis.

These experiments show that cp6396 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### 30 Example 111

The following *C.pneumoniae* protein (PID 4376408) was expressed <SEQ ID 221; cp6408>:

```

35      1  MNTSLKRPLK SHFDVVGSLF RPEHLKKTRE SLKEGSISLD QLMQIEDIAI
      51  QDLIKKQKAA GLSFITDGEF RRATWHYDFM WGFHGVGHHR ATEGVFFDGE
     101  RAMIDDTYLT DKISVSHHPF VDHFKEVKAL EDEFTTAKQT LPAPAQFLKQ
     151  MIFPNNIEVT RKFYPTNQEL IEDIVAGYRK VIRDLVDAGC RYLQLDDCTR
     201  GGLVDPRVCS WYGIDEKGLQ DLIQQYLLIN NLVIADRPDD LVVNLHVCRG
     251  NYHSKFASG SYDFIAKPLF EQTNVDGYL EFDHERSGDF SPLTFISGEK
     301  TVCLGLVTSK TPTLENKDEV IARIHQAADY LPLERLSLSP QCGFASCEIG
     351  NKLTEEEQWA KVALVKEISE EVWK*

```

40 The cp6408 nucleotide sequence <SEQ ID 222> is:

```

45      1  ATGAATACTT CACTAAAAAG ACCTCTGAAA TCTCATTTTG ATGTTGTGCG
      51  TAGTTTTTTG CGTCCTGAGC ATTTAAAAAA AACTAGAGAA AGCCTTAAAG
     101  AAGGCTCTAT TTCTCTAGAT CAACTCATGC AAATTGAGGA TATCGCTATC
     151  CAAGATTTGA TCAAAAAACA AAAAGCAGCA GGTCTTCTTT TTATTACTGA
     201  TGGAGAATTC CGCAGAGCTA CGTGGCATTG CGACTTCATG TGGGGTTTTT
     251  ATGGCGTAGG TCACCACAGA GCTACAGAAG GAGTTTTCTT TGATGGAGAA
     301  CGCGCTATGA TCGATGATAC CTATCTGACA GACAAGATCT CTGTATCTCA
     351  CCACCCATTT GTGGATCACT TTAAATTTGT AAAAGCTCTA GAAGATGAAT
     401  TTACGACTGC AAAGCAAACCT CTTCTCTGCAC CGGCACAGTT TTTAAAGCAG
     451  ATGATCTTCC CTAATAATAT AGAGGTCACA CGTAAATTCT ATCCTACAAA
     501  TCAGGAGCTA ATTGAAGATA TTGTTGCAGG TTATCGTAAA GTCATTCGCG
     551  ATCTTTTATGA TGCTGGCTGC CGCTATCTCC AATTAGATGA CTGACTCGCG
     601  GGAGGTTTAG TAGACCCTCG AGTCTGTTTC TGGTATGGTA TCGATGAAAA
     651  AGGTCTTCAA GATCTGATTC AACAAATATC TCTGATTAAT AATCTTGTA
     701  TTGCAGATCG TCCCGATGAT CTAGTCGTTA ATTTACATGT ATGCCGTGGG

```

-142-

201 CAAGGCTCCA CATTTAGATC CTGAAATCTA TAAACTTGGC ATTCCAATTC  
 251 TAGCTATTG CTATGGCATG CAGCTTATGG CTAGAGATTT TGGAGGGACT  
 301 GTAAGCCCTG GTGTAGGAGA ATTTGGATAT ACGCCCATCC ATCTGTATCC  
 351 TTGTGAGCTC TTCAAACACA TCGTCGACTG CGAATCTCTA GACACAGAGA  
 401 TTCGGATGAG CCATCGGGAT CATGTTACGA CAATTCCTGA AGGATTTAAT  
 451 GTAATCGCAT CCACCTCACA ATGCTCGATC TCAGGAATAG AAAATACCAA  
 501 ACAACGGTTG TACGGGCTGC AATTTTCATCC CGAGGTTTCT GACTCCACTC  
 551 CAACGGGAAA TAAGATTCTA GAAACTTTTG TTCAAGAGAT CTGTTCTGCT  
 601 CCCACACTAT GGAATCCCTT GTATATTCAG CAAGACCTTG TAAGTAAAT  
 651 TCAAGATACC GTTATTGAAG TATTTGATGA AGTCGCTCAG TCATTAGACG  
 701 TACAATGGTT AGCTCAAGGA ACCATCTACT CAGATGTTAT TGAGTCTCTA  
 751 CGCTCTGGAC ATGCCCTCGA AGTAATAAAA TCACATCATA ATGTAGGGGG  
 801 GCTTCCAAAA AATCTTAAGC TGAAGTTAGT CGAGCCCTTA CGTTATTTAT  
 851 TTAAAGATGA AGTTCGAATT TTAGGAGAAG CCCTAGGACT TTCTAGCTAT  
 901 CTCTTGGACA GGCATCCTTT TCCTGGACCT GGCTTGACAA TTCGTGTGAT  
 951 TGGAGAGATC CTTCTGAAT ATCTAGCCAT TTTACGACGG GCGGACCTCA  
 1001 TCTTTATAGA AGAGCTTAGG AAAGCAAAAC TCTACGATAA AATAAGCCAA  
 1051 GCCTTTGCTC TATTTCTTCC TATAAATCA GTATCTGTAA AAGGAGATTG  
 1101 TAGAAGCTAT GGTATATCCA TAGCATTACG TGCTGTAGAA TCTACAGATT  
 1151 TCATGACAGG ACGATGGGCC TACCTTCCAT GCGATGTTCT CAGTTCTTGC  
 1201 TCATCGCGAA TTATTAATGA AATACCCGAG GTAAGCCGAG TGGTCTATGA  
 1251 TATTTCTGAC AAGCCACCAG CAACTATAGA ATGGGAATAG

The PSORT algorithm predicts cytoplasm (0.0481).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 114A) and also as  
 a his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used  
 in a Western blot (Figure 114B) and for FACS analysis.

These experiments show that cp6440 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 115

The following *C.pneumoniae* protein (PID 4376475) was expressed <SEQ ID 229; cp6475>:

1 MNTYTFSP TL QKSFSFLLE KLDSYFFFGG TRTQILVITP TNIRLAAKKR  
 51 GCKVSTIEKI IKILSFILLP LVIIAFILRY FLHKKFDKQF LCIPKVISNE  
 101 DEALLGSRPQ AVEKAVREIS PAFFSIPRKY QLIRIDTPKD DAPSILFFPIG  
 151 IEIILKDL CI DTLKQSNLFL KREMDFLGHP EEKALFDSIC SIEKDQEWMS  
 201 LESKKLLITH FLKYL FVSGI EQLNPGFNPE NGRGYFSEIS TAKIHFHQHG  
 251 RYGPISRSGP IMKEI\*

The cp6475 nucleotide sequence <SEQ ID 230> is:

1 ATGAATACCT ATACCTTCTC TCCTACACTT CAGAAAAGCT TCAGCCTATT  
 51 TCTTTTAGAA AAATTAGACT CTTACTTTT CTTTGGAGGG ACTCGTACAC  
 101 AAATCTTAGT CATCACACCA ACCAATATTA GATTAGCAGC TAAAAAAGA  
 151 GGGTGTAAAG TTTCTACTAT AGAAAAGATA ATCAAGATCC TCTCTTTTAT  
 201 CCTGCTGCCC CTAGTTATCA TTGCCTTTAT ACTTCGCTAT TTCTTACATA  
 251 AGAAATTCGA TAAACAGTTC TTGTGTATCC CAAAAGTCAT TTCTAACGAA  
 301 GACGAAGCTC TTCTTGGATC TAGACCACAA GCAGTTGAAA AAGCAGTTCG  
 351 AGAAATATCT CCAGCCTTCT TCTCTATACC AAGAAAATAC CAACTTATTA  
 401 GAATCGACAC TCCTAAAGAT GACGCTCCCT CAATCCTTTT CCCTATAGGC  
 451 ATAGAGATCA TTCTCAAAGA TTTATGTATT GATACACTCA AGCAATCTAA  
 501 TCTTTTCTCT AAAAGAGAAA TGGATTTCTT AGGTCATCCA GAAGAAAAG  
 551 CATTATTCGA CTCGATATGT TCTATAGAAA AAGATCAAGA ATGGATGAGC  
 601 TTGGAAAGTA AAAAATTTT AATCACGCAC TTCCTAAAGT ATCTCTTTGT  
 651 CTCTGGAATC GAACAACATA ATCCAGGCTT TAACCCAGAG AATGGGCGTG  
 701 GGTATTTTTC AGAAATAAGT ACAGCAAAGA TCCATTTTCA TCAGCACGGT  
 751 CGATATGGGC CAATCCGTTT TTCGGGACCC ATCATGAAGG AAATATAA

The PSORT algorithm predicts inner membrane (0.5373).

**Example 113**

The following *C.pneumoniae* protein (PID 4376439) was expressed <SEQ ID 225; cp6439>:

```

1  MSYDTLFLKNL EKEDSVHKIC NEIFALVPRL NTIACTEAII KNLPKADIV
51 HLPGTITPQL AWILGVKNGF LKWSYNSWTN HRLLSPKNPH KOYSNIFRNF
101 QDICHEKDPD LSVLQYNILN YDFNSFDRVM ATVQGHRFPP GGIQNEEDLL
151 LIFNNYLQQC LDDTIVYTEV QQNIRLAHVL YPSLPEKHAR MKFYQILYRA
201 SQTFSKHGIT LRFLNCFNKT FAPQINTQEP AQEAVQWLQE VDSTFPGLFV
251 GIQSAGESA PGACPKRLAS GYRNAYDSGF GCEAHAGEGI ETRTIFSSAK
301 VNPEGLIEIT RVTFSSLKRK QPSSLPIRVT CQLG*

```

The cp6439 nucleotide sequence <SEQ ID 226> is:

```

1  ATGTCTTATG ATACGTTATT CAAGAATCTT GAAAAGGAAG ATTCTGTACA
51 TAAGATATGC AATGAGATCT TTGCATTAGT ACCACGACTC AATACAATCG
101 CTTGCACCGA AGCTATCATC AAAAACCTCC CCAAAGCAGA TATCCATGTA
151 CACCTTCC TG GACCATAAC ACCTCAATTA GCTTGGATT TAGGTGTGAA
15  201 AAATGGGTTT TTAATGGT CTTATAATTC TTGGACCAAT CATCGATTAC
251 TTTCTCCTAA GAATCCTCAT AAACAATACT CCAATATTTT CCGAACTTT
301 CAAGATATCT GTCACGAAAA GGATCCGGAT TTAAGTGTAT TACAATATAA
351 TATCTTAAAT TACGATTTTA ATAGCTTTGA TAGAGTGATG GCTACAGTAC
401 AAGGACATCG CTTTCTCCTT GGAGGAATCC AAAATGAAGA AGACCTTCTT
20  451 CTCATTTTCA ATAACATCTT CCAGCAATGT CTGGACGATA CTATCGTGTA
501 TACTGAAGTA CAACAAAATA TCCGCCTTGC CCATGTTTTG TATCCTTCAT
551 TACCTGAAAA GCACGCGCGT ATGAAGTTT ATCAAATCTT GTATCGTGCT
601 TCGCAAACGT TTTCAAAACA CGGGATTACT TTACGATTTT TAAACTGCTT
651 CAATAAAACA TTTGCTCCAC AAATAAACAC ACAAGAACCT GCCCAAGAG
25  701 CTGTTCATG GCTCCAAGAG GTTGATTCTA CATTTCTTGG TCTATTTGTA
751 GGGATACAAT CCGCAGGATC AGAATCTGCG CCCGAGCCT GTCCTAAGCG
801 ATTAGCTTCT GGATATAGAA ATGCTTATGA CTCAGGTTT GGTGTGTAAG
851 CTCATGCTGG AGAAGGCATA GAGACCCGGA CTATTTTTC GTCAGCTAAG
901 GTAAATCCAG AGGGATTGAT CGAGATAACC CGAGTGACTT TCTCGTCTCT
30  951 TAAACGAAAA CAGCCATCTA GTTTACCCAT AAGAGTTACT TGCCAGTTAG
1001 GATAA

```

The PSORT algorithm predicts cytoplasm (0.1628).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 113A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 113B) and for FACS analysis.

These experiments show that cp6439 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

**Example 114**

The following *C.pneumoniae* protein (PID 4376440) was expressed <SEQ ID 227; cp6440>:

```

1  LQSARRHLNT IFILDFGSQY TYVLAKQVRK LFVYCEVLPW NISVQCLKER
51 APLGIILSGG PHSVYENKAP HLDPEIYKLG IPILAICYGM QLMARDFGGT
101 VSPGVGEFGY TPIHLYPCEL FKHIVDCESL DTEIRMSHRD HVTTIPEGFN
151 VIASTSQCSI SGIENTKQRL YGLQFHPEVS DSTPTGNKIL ETFVQEICSA
201 PTLWNPLYIQ QDLVSKIQDT VIEVFDEVAQ SLDVQWLAQG TIYSDVIESS
45  251 RSGHASEVIK SHHNVGGLPK NLKLKLVEPL RYLFKDEVRI LGEALGLSSY
301 LLDRHPPFGP GLTIRVIGEI LPEYLAILRR ADLIFIEELR KAKLYDKISQ
351 AFALFLPIKS VSVKGDERSY GYTIALRAVE STDFMTGRWA YLPCDVLSSC
401 SSRIINEIPE VSRVVDISD KPPATIEWE*

```

The cp6440 nucleotide sequence <SEQ ID 228> is:

```

1  TTGCAGAGTG CAAGGAGACA TTTGAACACC ATATTTATTC TAGATTTTGG
51 ATCTCAATAT ACTTATGTAT TAGCAAAGCA AGTGCAGGAG TTATTTGTAT
101 ATTGCGAAGT TCTTCCCTGG AATATCTCTG TGCAATGTTT AAAAGAAAGA
151 GCGCCTTTGG GGATCATCTC CTCAGGAGGT CCTCACTCTG TCTATGAAAA

```

351 ALLVVRKIQFR GAIKSAYFEK LTBIEKELRS LQDVIKSLEL ELIHKIKDIV  
401 TEET\*

The cp6486 nucleotide sequence <SEQ ID 234> is:

```

5      1  GTGGTGGTTG TCGCTTTATT TATCCTTGGG ATTTTCTTTT TATCTGGTTC
      51  TCTTGCATTC CTTGTTTCATA CGTCTTGCGG AGTTCTTTTA GGAGCGGCGC
     101  TTCCCATACT TTGCATAGGT CTTGTTTAT TGGCTGTAGC TCTTATTGTT
     151  TTCTTATGTC ACAAACACAA GACTCGTCAA GATTTAGATT ATTATGATCA
     201  AGATTTAGAT TCTTTGGTGA TTCATAAGAA AGAGATCCCC AATGACATCT
     251  CTGAGTTGCG GGTAAACATT GAAAAGTTGC AAAATCTGTT TCAGTTCCAT
    10  301  ACGAAAGATT TCTCTGATCT AAGCCAAGAG CTTCAGGGTA AATTTATCAA
     351  TTGCATGGAG AAATGCCTAA CTTTAGAAGA CGAAGTGACT AAATTTCTTA
     401  TTGTTTCGAGA TAGATTTTTA GAAACCAGAA GAAATTTTAC CACTTTTGGA
     451  GAACAGGTTA AAGGGATCCA AAGCAATATT TTTGATTTGC ATGAGGAAAA
    15  501  TTCTTCATTA TATTAGAAT TGTATAGGCT TAGGAAAGAC CTCCAAGTTC
     551  TATTAAATTT TTTTCTGCTC CCCCCAGGTA TACTCAAGGT AGATTATGAT
     601  GAAATTGAGG CTATCAAAGG TCTGTTTATA AGATTAACCT CTAGATTAGA
     651  TAAGCTTGAT GTGAAAGCTC AGGAACGTAA GAAGTTCATT AATGAAATGA
     701  GTAGGGAATT TAAAGAAGTA GAGAAAGCTT TTGATATTGT CGATAGGGCA
     751  ACAAAAAGC TTATGGATAG AGCCAAGAAA GAAAGTCCGG CACGCTTTT
    20  801  CATGGGTAGA ACTGAGTCTC TCTTAGAAAT GAAAAAAAT GAAGAAGCCC
     851  TTAAAAATCA GGGGCTAGAT CCTGAAAATC TTTCCCATCC TGAACTTTT
     901  AGTCCGTATC AACAGCTTTT AATTTTGAAT TATTAAATA GCGAAATAGT
     951  TCTGCATCAT TATGAGTTCC TTATTTCTGG AACAGTAACT TCTGGCCTAA
    25 1001  CTCTTGAAGA ATGTGAAAAT CGAATGAGGG CGGCTTCTAC TGGGTTGAAC
     1051  GCCCTTCTGG TGCCTAAGCT CCAGTTCAGA GGTGCTATAA AATCTGCGTA
     1101  TTTTGAAAAA CTCACAGAGA TTGAAAAGA GTTACGATCA CTTCAAGACG
     1151  TAATAAAGTC ATTGGAACTA GAACTGATCC ATAAGATAAA AGATATAGTG
    1201  ACAGAAGAAA CTTAG

```

The PSORT algorithm predicts inner membrane (0.7474).

30 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 117A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 117B) and for FACS analysis.

These experiments show that cp6486 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### 35 Example 118

The following *C.pneumoniae* protein (PID 4376526) was expressed <SEQ ID 235; cp6526>:

```

40  1  MSPFKKIVNR LLCYISFQKE SRTLPIIIRE PRMTTKSLGS FNSVISKNKI
      51  HFISLGCSR N LVDSEVMLGI LLKAGYESTN EIEDADYLIL NTCAFLKSAR
     101  DEAKDYLDHL IDVKKENAKI IVTGCMTSNH KDELKPWMSH IHYLLGSGDV
     151  ENILSAIESR ESGEKISAKS YIEMGEVPRQ LSTPKHYAYL KVAEGCRKRC
     201  AFCIIPSIKG KLRSKPLDQI LKEFRILVNK SVKEIILIAQ DLGDYGKDLS
     251  TDRSSQLESL LHELLKEPGD YWLRMLYLYP DEVSDGIIDL MQSNPKLLPY
     301  VDIPLQHIND RILKQMRRTT SREQILGFLE KLRKVQVY IRSSVIVGFP
     351  GETQE EFQEL ADFIGEWID NLGIFLYSQE ANTPAAELPD QIPEKVKESR
    45  401  LKILSQIQKR NVDKHNQKLI GEKIEAVIDN YHPETNLLLT ARFYGQAPEV
     451  DPCIIVNEAK LVSHFGERC F IEITGTAGYD LVGRVVKKSQ NQALLKTSKA
     501  *

```

The cp6526 nucleotide sequence <SEQ ID 236> is:

```

50  1  ATGAGTCCTT TTAAGAAAAT AGTAAATCGC TTACTATGCT ATATTTCTTT
      51  TCAAAAAGAA TCAAGAACTC TCCAATCAT TATTAGAGAA CCTAGGATGA
     101  CAACAAAAG TTTAGGATCT TTCAATTCAG TTATTCCAA AAATAAAATT
     151  CATTTTATTA GTTTGGGATG CTCTCGGAAC CTGTAGATA GCGAAGTCAT
     201  GCTAGGCATT CTTCTTAAGG CAGGTTACGA GTCTACTAAT GAAATTGAAG
     251  ATGCTGACTA TTTAATTTTA AATACCTGTG CGTTTTTAAA AAGTGCTAGA
    55  301  GATGAAGCTA AAGATTATCT AGACCATCTA ATTGATGTAA AAAAAGAGAA

```

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 115A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 115B) and for FACS analysis.

These experiments show that cp6475 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 116

The following *C.pneumoniae* protein (PID 4376482) was expressed <SEQ ID 231; cp6482>:

```

1  MLVELEALKR EFAHLKDQKP TSDQEITSLY QCLDHLEFVL LGLGQDKFLK
51  ATEDEDVLF E SQKAIDAWNA LLTKARDVLG LGDIGAIYQT IEFLGAYLSK
101 VNRRAFCIAS EIHFLLKTAIR DLNAYLLDF RWPLCKIEEF VDWGNDCEVEI
151 AKRKLCTFEK ETKELNESLL REEHAMEKCS IQDLQRKLSL IIEHLHDVSL
201 FCFSKTPSQE EYQKDCLYQS RLRVLLLLYE YTLCKTSTD FQEQRARKEE
251 FIREKPSLLE LERGIKQTK E LEFAIAKSKL ERGCLVMRKY EAAAKHSLDS
301 MFEEETVKSP RKDTE*

```

The cp6482 nucleotide sequence <SEQ ID 232> is:

```

1  ATGCTAGTAG AGTTAGAGGC TCTTAAAAGA GAGTTGCGC ATTTAAAAGA
51  CCAGAAGCCG ACAAGTGACC AAGAGATCAC TTCACTTTAT CAATGTTTGG
101 ATCATCTTGA ATTCGTTTTA CTCGGGCTGG GCCAGGACAA ATTTTAAAG
151 GCTACGGAAG ATGAAGATGT GCTTTTGTAG TCTCAAAAAG CAATCGATGC
201 GTGGAATGCT TTATTGACAA AAGCCAGAGA TGTTTTAGGT CTTGGGGACA
251 TAGGTGCTAT CTATCAGACT ATAGAATTCT TGGGTGCCTA TTTATCAAAA
301 GTGAATCGGA GGGCTTTTTG TATTGCTTCG GAGATACATT TTCTAAAAAC
351 AGCAATCCGA GATTTGAATG CATATTACCT GTTAGATTTT AGATGGCCTC
401 TTTGCAAGAT AGAAGAGTTT GTGGATTGGG GGAATGATTG TGTGAAATA
25  451 GCAAAGAGGA AGCTATGCAC TTTTGAAAAA GAAACCAAGG AGCTCAATGA
501 GAGCCTTCTT AGAGAGGAGC ATGCGATGGA GAAATGCTCG ATTCAAGATC
551 TGCAAAGGAA ACTTAGCGAC ATTATTATTG AATTGCATGA TGTTCCTCTT
601 TTTTGTTTTT CTAAGACTCC CAGTCAAGAG GAGTATCAAA AGGATGTTT
651 GTATCAATCA CGATTGAGGT ACTTATTGTT GCTGTATGAG TATACATTGT
30  701 TATGTAAGAC ATCCACAGAT TTTCAAGAGC AGGCTAGGGC TAAAGAGGAG
751 TTCATTAGGG AGAAATTCAG CCTTCTAGAG CTCGAAAAGG GAATAAACA
801 AACTAAAGAG CTTGAGTTTG CAATTGCTAA AAGTAAGTTA GAACGGGGCT
851 GTTTAGTTAT GAGGAAGTAT GAAGTCGCCG CTAAACATAG TTTAGATTCT
901 ATGTTTGAAG AAGAACTGT GAAGTCGCCG CGGAAAGACA CAGAATAA

```

The PSORT algorithm predicts cytoplasm (0.4607).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 116A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 116B) and for FACS analysis.

These experiments show that cp6482 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 117

The following *C.pneumoniae* protein (PID 4376486) was expressed <SEQ ID 233; cp6486>:

```

1  VVVVALFILG IFFLSGSLAF LVHTSCGVLL GAALPILCIG LVLLAVALIV
51  FLCHKHKTRQ DLDYYDQDL D SLVIHKKIEP NDISELRVTF EKLQNLFPQH
45  101 TKDFSLSQ E LQKFINCME KWLTLDEV T KFLIVRDRFL ETRRNFITFG
151 EQVKGIQSN I FDLHEEKSSL YLELYRLRK D LQVLLNFFLL PPGILKVDYD
201 EIEAIKGLF I RLTSRLDKL D VKAQERKKF I NEMSREFKEV EKAFFDIVDRA
251 TKKLMDRAK K ESPARLFMGR TESLLEMKN E EALKNQGLD PENLSHPELF
301 SPYQQLLIL N YLNSEIVLH H YEFLISGTV T SGLTLEECEN RMRAASTGLN

```

The PSORT algorithm predicts cytoplasm (0.1668).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 119A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 119B) and for FACS analysis.

- 5 These experiments show that cp6528 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 120

The following *C.pneumoniae* protein (PID 4376627) was expressed <SEQ ID 239; cp6627>:

```

10      1  MKCSPLTLVP HIFLKNDEC HRSCSLKIRT IARLILGLVL ALVSALSFVF
      51  LAAPISYAIG GTLALAAIVI LIITLVVALL AKSKVLPIPN ELQKIYNRY
     101  PKEVFYFVKI HSLTVNELKI FINCWKSGTD LPPNLHKKAE AFGIDILKSI
     151  DLTLPPEFEE ILLQNCPLYW LSHFIDKTES VAGEIGLNKT QKVYGLLGPL
     201  AFHKGYTTIF HSYTRPLLT ISESQYKFLY SKASKNQWDS PSVKKTCEEI
     251  FKELPHNMIF RKDVQGISQF LFLFFSHGIT WEQAQMIQLI NPDNWKMLCQ
15     301  FDKAGGHCSM ATFGGFLNTE TNMFDPVSSN YEPTVNFMTW KELKVLLEKV
     351  KESPMHPASA LVQKICVNTT HHQNLKRWQ FVRNTSSQWT SSLPQYAFHA
     401  QTYKLEKKIE SSLPIRSSL*
```

The cp6627 nucleotide sequence <SEQ ID 240> is:

```

20      1  ATGAAGTGTA GTCCTTTAAC ACTAGTTCCC CATATATTTT TAAAAAATGA
     51  CTGCGAATGT CATAGATCTT GTTCTTTAAA AATTAGGACA ATTGCCCCGAC
     101  TCATTCTTGG GCTTGTTCTA GCTCTTGTTA GCGCACTTTC TTTTGTTTTC
     151  CTTGCTGCGC CGATTAGCTA TGCTATTGGA GGAACCTTAG CTTTAGCCGC
     201  TATCGTAATC TTGATTATAA CGCTAGTCGT AGCACTGCTA GCTAAATCAA
     251  AGGTTCTGCC CATCCCCAAC GAACTTCAGA AGATTATTTA CAATCGCTAT
25     301  CCTAAAGAAG TCTTTTATTT CGTGAAAACA CACTCCCTGA CTGTTAACGA
     351  ATTAAAAATA TTTATTAATT GCTGGAAAAG CCGTACAGAC CTGCCTCCGA
     401  ATTTACATAA AAAAGCAGAG GCTTTCGGGA TCGATATTCT AAAATCTATA
     451  GATTTAACCC TGTTCACAGA GTTCGAAGAG ATTCTTCTTC AAAACTGCCC
     501  GTTATACTGG CTCTCCCAT TATAGACAA AACTGAATCT GTTGCTGGGG
30     551  AAATCGGATT AAATAAAACA CAAAAGTTT ATGGTTTACT TGGGCCCTTA
     601  GCGTTTCATA AAGGATATAC AACTATTTTC CACTCTTATA CACGCCCTCT
     651  ACTAACATTA ATCTCAGAAT CACAGTATAA GTTCCTATAT AGTAAAGCGT
     701  CTAAGAATCA ATGGGATTCT CCTTCTGTGA AAAAACCCTG CGAAGAAATA
     751  TTCAAGGAAC TCCCCACAA TATGATTTTC CGGAAGGATG TTCAAGGAAT
35     801  CTCACAATTC TTATTTCTTT TCTTTTCTCA TGGTATCACT TGGGAACAGG
     851  CTCAGATGAT TCAACTTATA AATCCTGATA ATTGGAATAT GTTGTGTCAG
     901  TTTGATAAAG CAGGAGGCCA CTGTTCCATG GCAACATTG GAGGCTTTTT
     951  GAATACTGAA ACAAATATGT TCGATCCAGT ATCCTCTAAC TATGAACCTA
40    1001  CAGTGAACCT CATGACGTGG AAAGAATTGA AGGTTTTACT AGAGAAAGTA
     1051  AAAGAAAGTC CTATGCACCC AGCGAGTGCT CTTGTTTACA AGATATGCGT
     1101  AAATACAACG CACCATCAAA ATCTGTTAAA ACGATGGCAA TTTGTTTCGT
     1151  ATACGAGTTC ACAATGGACA TCAAGCTTAC CTCAGTATGC TTTCCACGCC
     1201  CAAACCTACA AACTAGAGAA AAAAATAGAA AGCAGTCTCC CTATACGATC
     1251  TTCCCTATAA
```

- 45 The PSORT algorithm predicts inner membrane (0.7198).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 120A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 120B) and for FACS analysis.

- 50 These experiments show that cp6627 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.



-145-

351 CGCTAAAATT ATTGTAAC TG GATGCATGAC TTCCAACCAC AAAGATGAGC  
 401 TTAAACCCTG GATGTCACAC ATCCATTACC TACTAGGTTT TGGGGATGTT  
 451 GAGAAATATC TTTC TGCTAT TGAGTCTCGT GAATCTGGAG AAAAAATCTC  
 501 TGCAAAGAGT TACATTGAGA TGGGAGAAGT TCCAAGACAG CTTTCCACAC  
 551 CAAACACTA TGCCTATTTA AAAGTTGCTG AGGGCTGTAG AAAACGTTGT  
 601 GCTTTTGTG TATTCCCTTC CATTAAAGGA AAGCTCCGCA GCAAACCTCT  
 651 GGATCAAATT CTTAAAGAAT TCCGCATCCT TGTAAACAAG AGTGTGAAAG  
 701 AGATTATATT GATAGCTCAA GACCTAGGAG ATTATGGAAA GGATCTCTCT  
 751 ACAGACCGCA GTTCGCAGCT AGAATCACTA TTACATGAGT TACTGAAAGA  
 801 GCCTGGTGAT TATTGGCTGC GGATGTTGTA TTTATATCCT GATGAAGTGA  
 851 GTGATGGCAT TATAGATCTT ATGCAATCTA ATCCCAAAC TCTTCCCTAT  
 901 GTAGATATTC CCTTACAGCA CATTAAACGAC CGTATTTTAA AGCAAATGCG  
 951 AAGAACGACT TCTAGGGAGC AAATCCTAGG ATTCCTAGAA AAATTACGTG  
 1001 CCAAGGTTCC TCAGGTCTAT ATCCGTTCTT CTGTTATTGT GGGTTTCCCC  
 1051 GGTGAAACTC AGGAAGAATT CCAGGAGTTA GCTGATTTTA TTGGTGAGGG  
 1101 TTGGATTGAT AATCTCGGAA TTTTCTGTG CTCTCAAGAA GCGAATACCC  
 1151 CGGCAGCAGA ACTCCCTGAC CAGATACCAG AAAAAAGTTAA AGAATCAGAG  
 1201 TTGAAAATTC TATCTCAAAAT TCAGAAACGC AATGTGGATA AACATAATCA  
 1251 GAAGTCATT GGGGAAAAAA TAGAAGCAGT TATTGATAAC TATCATCCTG  
 1301 AAACGAATCT TTTACTCACT GCAAGGTTCT ATGGACAAGC TCCTGAAGTG  
 1351 GACCCTTGTA TTATTGTAAA TGAGGCGAAG CTTGTTTCTC ATTTTGGAGA  
 1401 AAGATGCTTT ATAGAAATCA CAGGGACTGC TGGTTACGAC CTTGTAGGGC  
 1451 GTGTTGTAAA AAAATCTCAG AACCAAGCTT TGCTAAAAAC TAGCAAAGCT  
 1501 TAG

25 The PSORT algorithm predicts cytoplasm (0.1296).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 118A) and also as a his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 118B) and for FACS analysis.

30 These experiments show that cp6526 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 119

The following *C.pneumoniae* protein (PID 4376528) was expressed <SEQ ID 237; cp6528>:

1 MKNNINNEC YFKLDSTVDG DLLAANLKTG DTQAQGISST ETFSVQGNAT  
 35 51 FKDQVSATGL TSGTFYNLNA QNFTSSQISI DFKNNRLSNC ALPKEDCDPV  
 101 PANYVRSPEY FFC SKPLIGD FDFNSGESYL PLTGSEYTLY QSRNVNSIFR  
 151 FIGWKQSTRE LTVGGNTAIQ FLAAGTYIVS FTVGKRWGWN NGWGGATYIN  
 201 NGLGQVQCES TIYSGGYAT IGTLTSTIYR ASVDVAPNPN DPNASDRYRA  
 251 GIFYLSNGGS SAGIGNYSFS LLYPPDRG\*

The cp6528 nucleotide sequence <SEQ ID 238> is:

40 1 ATGAAAAACA ATATTAATAA TAATGAGTGC TATTTTAAAT TAGACTCAAC  
 51 TG TAGATGGT GATTTGTTAG CAGCCAATCT CAAGACCTTT GATACACAGG  
 101 CCCAAGGAAT CTCATCGACT GAAACATTTT CTGTTTCAGGG GAATGCAACA  
 151 TTTAAAGATC AAGTTTCAGC AACTGGATTA ACTTCAGGAA CTACTTATAA  
 201 TTTAAATGCA CAAACTTTTA CTTCTCCTCA AATCTCTATA GATTTTAAAA  
 45 251 ATAATCGTCT GAGTAATTGT GCATTGCCAA AAGAAGACTG CGATCCGGTG  
 301 CCAGCGAATT ATGTTTCGTT TCCCGAATAT TTTTCTGTG CCAAGCCTCT  
 351 GATCGGAGAT TTTGATTTTA ACTCAGGGGA ATCTTATTTG CCTCTGACTG  
 401 GTTCGGAATA TACTCTATAT CAGTCACGTA ATGTAAATAG TATATTTCTG  
 451 TTTATAGGAT GGAAGCAAAG TACACGAGAA TTAACGTGAG GGGGAAATAC  
 50 501 TGCGATACAA TTTCTTGCAG CAGGAACCTA TATCGTTTCA TTTACTGTTG  
 551 GTAAACGGTG GGGATGGAAT AATGGTTGGG GAGGAGCCAT TTATATCAAT  
 601 AATGGTTTAG GACAAGTCCA ATGTGAAAGC ACGATTTATA GTGGTGGAGG  
 651 GTATGCAACA ATAGGTACAC TGGGGACCTC AATATATAGA GCCTCTGTAG  
 701 ATGTAGCTCC TAATCCTAAT GATCCGAATG CTTCCGATCG CTATAGAGCG  
 55 751 GGTATTTTCT ATCTCAGTAA CGGTGGTTCT AGTGCAGGTA TAGGGAATTA  
 801 CTCCTTTTCT CTTCTCTATT ATCCGGACGA TAGAGGGTAG

351 TEEEQWKKIA FVKEIAKEIW G\*

The cp6732 nucleotide sequence <SEQ ID 244> is:

```

      1  ATGGAAATGA  TGAGCCCATT  CCAACAACCT  GAGCAATGTC  ATTTTGATGT
      5  51  TGTGGGAAGT  TTCTTACGTC  CTGAAAGTCT  TACACGAGCA  CGCTCTGATT
      101  TTGAAGAAGG  AAGAATTGTC  TATGAGCAGA  TGCAGATTGT  CGAAGATGCT
      151  GCTATTCGTA  ATCTCATAAA  AAAGCAAACA  GAAGCAGGTC  TTATCTTTT
      201  TACTGATGGG  GAATCCCGTA  GGTATAGTTG  GGATTTGAC  TTTATGTGGG
      251  GATTCCATGG  CGTGGATCGT  CGCAGGGACT  CTAATGACCC  TGAAATTGGA
      301  GTGTATCTTA  AAGATAAAAT  CTCCGTATCA  AAACATCCGT  TTATAGAACA
      351  TTTTCGAGTTT  GTCAAACTT  TTGAGAAGGG  AAATGCAAAA  GCAAAACAAA
      401  CGATTCCCTC  TCCATCACAA  TTTTTCATG  AGATGATTTT  TGCTCCTAAT
      451  CTGAAAAATA  CTCGGAAGTT  TTATCCTACG  AATCAAGAGC  TAATTGATGA
      501  TATTGTCTTT  TATTATCGCC  AAGTCATCCA  AGATCTTTAT  GCTGCAGGTT
      551  GTCGTAATTT  GCAGTTGGAC  GATTGTGCTT  GGTGTCGCCT  CTTGGATATA
      601  CGAGCGCCTT  CTTGGTATGG  TGTGATTCT  CATGACAGGT  TGCAGGAAAT
      651  TTTAGAACAG  TTTTATGGA  TCCATAATTT  AGTGATGAAG  GATAGACCCG
      701  AGGATCTTTT  TGTAAGTCTG  CATGTCTGTC  GTGGTGATTA  TCAGGCCGAG
      751  TTTTCTCTA  GACGAGCTTA  TGATTCTATA  GAGGAGCCTT  TATTGCTAA
      801  GACCGATGTG  GATAGTTATC  ACTATTATTG  GGCTCTTGAT  GATAAGTATT
      851  CAGGAGGTGC  TGAGCCTTTA  GCTTACGTCT  CTGGAGAGAA  ACACGTCTGC
      901  TTGGGATTGA  TCTCCAGCAA  CCATTCTTGT  ATTGAAGATC  GAGATGCTGT
      951  GGTTCCTCGT  ATTTATGAAG  CTGCGAGCTA  CATCCCTT  GAGAGACTTT
     1001  CTTTGAGCCC  GCAATGTGGG  TTTGCTCTT  GTGAGGGAGA  CCATAGAATG
     1051  ACTGAAGAAG  AACAGTGGA  GAAGATCGCC  TTTGTGAAAG  AGATTGCTAA
     1101  AGAGATCTGG  GGATAA
  
```

The PSORT algorithm predicts cytoplasm (0.2196).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 122A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 122B) and for FACS analysis.

30 These experiments show that cp6732 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 123

The following *C.pneumoniae* protein (PID 4376738) was expressed <SEQ ID 245; cp6738>:

```

      1  VWLRFLLLV  YDEKEKDVVV  VCNHSEPNIL  GLPPEAVSQL  IEELSDEGYS
      35  51  YLNVVRCDL  GETTVQQRLL  LNADEGRSMT  VVISELPEGH  PDIRNLQLAS
      101  ERIFVSREKE  AADAYASGCK  VVAFDDEHLP  WVSSHIAAYE  EIREKQEQTM
      151  QGSLTEEQLG  ALLCNTVSTE  KNLAFALDAV  IKQSVWFRFN  PDLFAYEREA
      201  LEASVTDALV  SYVSNLDMIP  YTSSQGIVIE  DSSIVRTSQE  HTLIVNCAAF
      251  DKLASQIEFL  CPSDVLPI  KDPLISDDED  EELNPKVSSA  ADSKDKT*
  
```

40 The cp6738 nucleotide sequence <SEQ ID 246> is:

```

      1  GTGTGGCTGC  GCTTTTTACT  TTTAGTGTCC  TATGATGAGA  AGGAGAAAGA
      51  CGTAGTTGTC  GTTTGTAATC  ATTCTGAACC  TAATATCCTC  GGCCTGCCTC
      101  CTGAAGCAGT  CTCTCAGCTT  ATTGAAGAGC  TTAGCGATGA  AGGCTATAGC
      151  TATCTGAATG  TAGTGCGTTG  TGATCTCTCC  GGGGAGACTA  CGGTTCAACA
      45  201  ACGTCTGCTA  TTGAATGCCG  ATGAAGGGAG  ATCTATGACG  GTGGTGATCT
      251  CAGAGCTTCC  TGAAGGGCAC  CCCGATATTC  GGAATTTGCA  GTTGGCATCC
      301  GAAAGAATTT  TTGTTTCTCG  TGAAAAAGAA  GCTGCTGATG  CCTATGCTTC
      351  AGGATGTAAA  GTGGTCGCTT  TCGATGATGA  GCATCTCCCT  TGGGTCTCCA
      401  GTCATATTGC  CTACGCGGAG  GAGATCAGAG  AGAAACAAGA  ACAACAATG
      451  CAAGGGTCTT  TAACTGAAGA  GCAGTTAGGA  GCACTCCTCT  GCAACACAGT
      501  CTCCACAGAG  AAAAATCTAG  CCTTTGCTCT  AGACGCCGTG  ATAAAACAGT
      551  CTGTGTGGAG  ATTCCGCAAT  CCGGATCTTT  TTGCTTATGA  GAGAGAAGCT
      601  CTAGAGGCTT  CAGTAACAGA  TGCTTTAGTA  TCTTACGTTT  CAAATTTAGA
      651  CATGATACCG  TACACAAGTT  CTCAGGGCAT  AGTCATAGAA  GATAGTAGTA
      701  TCGTCCGTAC  CTCTCAAGAG  CATACTCA  TTGTGAACTG  TGCAGCATT
  
```

**Example 121**

The following *C.pneumoniae* protein (PID 4376629) was expressed <SEQ ID 241; cp6629>:

```

      1 MSNITSPVIQ NNRSNYYFE LKNSTTIHIV ISAILLCGAL IAFLCVAAPV
      51 SYILSGALLG LGLLIALIGV ILGIKKITPM ISSKEQVFPQ ELVNRIRAHY
      101 PKFVSDFVSE AKPNLKDLS FIDLLNQLHS EVGSSTNYNV SEELQQKIDT
      151 FEGIARLKNE VRTASLKRLE SAASSRPLFP SLPKILQKVF PFFWLGEFIS
      201 AGSKVVELHR VKKIGGSLEE DLSDYIKPEM LPTYWLIPLD FRPTNSSILN
      251 LHTLVLARVL TRDVFOHLKY AALNGEWNLN HSDLNTMKQQ LFAKYHAAYQ
      301 SYKHLSPQSL QEDEFYNLLL CIFKHRYSWK QMSLIKTVP DLWENLCLLT
      10 351 LDHTGRPDQM EFASLIGTLY TQGLIHKESE AFLSSLTLLS LDQFKTIRRO
      401 STNIAMFLEN LATHNSTFRS LPPIIVHPLK RSVFSQPEED ESSLLIG*

```

The cp6629 nucleotide sequence <SEQ ID 242> is:

```

      1 ATGAGTAATA TAACCTCGCC AGTTATTCAA AATAATCGCT CTTGTAATTA
      51 TTATTTTGAA TTAAGAATT CAACCACTAT TCATATTGTT ATCAGTGCCA
      15 101 TCTTACTCTG CGGAGCTTTG ATAGCTTTCT TGTGTGTAGC AGCTCCTGTT
      151 TCCTATATTC TAAGTGGCGC ATTGTTAGGA TTAGGATTAT TAATAGCCTT
      201 GATTGGTGTG ATTTTAGGAA TAAAAAAAT CACGCCATG ATTTCATCAA
      251 AAGAACAAGT ATTCCCCAA GAACCTCGTA ATAGAATCAG GGCGCACTAT
      301 CCTAAATTTG TCTCTGATT TGTTCAGAA GCTAAACCAA ATCTTAAAGA
      20 351 TCTCATAAGT TTTATTGATC TTCTAAATCA ATTGCACTCT GAAGTTGGAT
      401 CATCTACAAA TTACAACGTA TCTGAAGAAC TACAACAGAA AATAGATACG
      451 TTCGAGGGTA TCGCACGCTT AAAAAATGAA GTCCGTA CTCTCTTAA
      501 AAGACTTGAA AGCGCTGCTT CTTCCCGTCC CCTCTTCCCC TCTTTACCAA
      551 AAATCTTACA AAAGGTATTT CCATTTTCTT GGTTAGGAGA GTTTATTCT
      25 601 GCAGGCAGCA AGGTTGTAGA GCTCCATCGA GTTAAGAAAA TTGGAGGCAG
      651 CCTCGAAGAA GACCTTAGTG ATTATATAAA ACCAGAGATG CTTCTACCT
      701 ATTGGTTGAT TCCTTTAGAT TTTAGACCAA CAAATTCCTC TATTCTAAAT
      751 CTACACACAT TAGTTTTAGC TAGAGTCTTA ACTCGTGATG TTTTCAACA
      801 TCTTAAGTAT GCAGCATTAA ATGGCGAGTG GAACCTGAAT CATAGTGATC
      30 851 TAAATACTAT GAAACAGCAG CTCTTTGCTA AATATCATGC GCGGTATCAA
      901 TCCTATAAAC ATCTATCTCA ACCCTCTCTT CAAGAGGATG AATCTATAA
      951 CCTGCTCTTG TGTATTTTTA AGCATAGGTA CTCGTGGAAG CAGATGTCCT
      1001 TAATAAAAAC AGTCCCGGCT GATTTATGGG AAAACCTCTG TTGCTTGACT
      1051 TTAGACCATA CAGGACGACC CCAAGACATG GAATTGCGCT CTCTAATTGG
      35 1101 TACTCTCTAC ACACAAGGCC TAATTCATAA AGAAAGCGAA GCATTTCCTT
      1151 CTTCAATGAC ACTCCTTAGT TTAGATCAGT TTAACACGAT CCGTCGTGAC
      1201 TCAACCAATA TAGCGATGTT CCTTGAGAAT TTAGCACTC ATAATTCAC
      1251 CTTTAGAAGC TTACCACCTA TAACAGTCCA TCCACTCAAG AGAAGCGTCT
      1301 TCTCCCAACC TGAAGAAGAC GAGTCCTCCC TGCTGATAGG TTAG

```

40 The PSORT algorithm predicts inner membrane (0.5776).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 121A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 121B) and for FACS analysis.

45 These experiments show that cp6629 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

**Example 122**

The following *C.pneumoniae* protein (PID 4376732) was expressed <SEQ ID 243; cp6732>:

```

      1 MEMMSPFQOP EQCHFDVVG S FLRPESLTRA RSDFEGRIV YEQMRVVEDA
      51 AIRNLIKQOT EAGLIFFTDG EFRYSWDFD FMWGFHGVDR RRDSNDPEIG
      50 101 VYLKDKISVS KHPFIEHFEP VKTFEKGNAK AKQTIPSPSQ FFHEMIFAPN
      151 LKNTRKFYPT NQELIDIVF YYRQVIQDLY AAGCRNLQLD DCAWCRLLDI
      201 RAPSWEYVDS HDRLQEILEQ FLWIHNLVMK DRPEDLFVSL HVCRGDYQAE
      251 FFSRRAYDSI EEPLFAKTDV DSYHYWALD DKYSGGAEPL AYVSGEKHVC
      301 LGLISSNHSC IEDRAVVSR IYEAASYIPL ERLSLSPQCG FASCEGDHRM

```

These experiments show that cp6739 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 125

The following *C.pneumoniae* protein (PID 4376741) was expressed <SEQ ID 249; cp6741>:

```

5      1  MASCLSAWFS  IVREHFYRAF  DFLSLPFCARI  TEFVLGVIKG  IPVVGHIIVG
      51  IEWLVSRYLE  SFVTKTPTFVS  DVVSLLKTEK  VAGRDHIARV  VETLKRQRVA
     101  VAPEDEDKVH  GKIPVHPFGG  IQPVEVLTLY  PEVQDATLGL  AFSKIRNRVR
     151  QAYLQAPRPK  LQKIYIIGND  MNPFEVDDFL  HLARLCNETQ  RLYPDATISL
     201  YLTASGGRNA  MDKKNRKLLS  DCELNPKIAC  LDFNQGDVVK  QATCDCWMVY
10     251  HGENDQGTLN  QIQEELEKSG  EETPWIVHGQ  KPLSQSLWDF  SPFSSLEMKG
     301  DKEKALEYSE  LEKEQLYSRL  VYVGERSSVL  SLGFGDSRSG  ILMDPKRVHA
     351  PLSEGHYCHS  YLADLENPGL  QKTILAAFLN  PKELSSTILQ  PISLNLILNS
     401  KTYLRQHFGF  FERMSRSDRN  VVVVVCDSWW  GTDWKEEPSF  QHFIMELECR
     451  GYSHFNIFAF  RSNSMCVEER  RILNESSQEK  AFTMIFCEDS  VSQGDIRCLH
15     501  LASEGMLCGK  ECVADVYTS  GCANFMEEV  LTLERESNLW  NRKHGLWKRE
     551  VRKQKQEAAL  DQDESEIYVC  NQLTAQQNFA  CS*

```

The cp6741 nucleotide sequence <SEQ ID 250> is:

```

      1  ATGGCTTCTT  GTTTATCTGC  CTGGTTTCTT  ATAGTTCGTG  AGCACTTTTA
      51  TCGAGCCTTT  GATTTTTCTT  TGCCGTTTTG  TGCTCGTATT  ACGGAATTTG
20    101  TATTAGGGGT  CATCAAGGGG  ATCCCTGTTG  TGGGTCACAT  TATTGTTGGG
     151  ATAGAGTGGC  TCGTTTCTAG  GTATTTAGAG  AGTTTCGTGA  CCAAGCCGAC
     201  ATTTGTCTCT  GATGTGGTGA  GTCTTCTGAA  AACAGAGAAA  GTTGCTGGTC
     251  GCGATCACAT  TGCTCGTGTA  GTGGAGACTT  TGAAGAGGCA  GAGAGTCGCT
30    301  GTGGCTCCTG  AAGATGAGGA  TAAGGTCCAT  GGAAGATTTC  CTGTGCATCC
     351  TTTCGGGGGA  ATCCAACCTG  TAGAAGTTCT  CACTCTCTAT  CCCGAAGTTC
     401  AAGATGCAAC  GTTAGGGCTT  GCCTTCTCTA  AAATTCGTAA  TCGTGTAAGA
     451  CAGGCGTATT  TGCAAGCTCC  ACGGCCAAAA  CTGCAGAAGA  TTTACATCAT
     501  AGGAAACGAT  ATGAATCCTT  TTGAAGTTGA  CGACTTCTTG  CATCTAGCCC
35    551  GTCTCTGTAA  TGAAACTCAA  AGACTCTATC  CTGACGTAC  GATTTCTCTA
     601  TATCTAACAG  CTTCTGGTGG  TCGCAATGCT  ATGGACAAAA  AGAATCGGAA
     651  GTTACTTAGT  GATTGCGAAC  TAAACCCCAA  GATTGCTTGT  TTGGACTTTA
40    701  ATCAGGGTGA  TGTAGTCAAA  CAAGCAACTT  GTGACTGTG  GATGGTGTAT
     751  CATGGGGAGA  ATGATCAAGG  TACGTTGAAT  CAGATTCAGG  AAGAGTTAGA
     801  AAAGTCAGGG  GAGGAAACCC  CTTGGATTCA  TGTGGGGCAA  AAGCCTCTTT
35    851  CACAATCCTT  GTGGGATTTT  TCTCCATTTT  CATCTTTGGA  GATGAAGGGA
     901  GATAAAGAGA  AAGCTCTAGA  GTACTCTGAA  TTAGAAAAAG  AACAGCTATA
     951  TTCTCGATTG  GTATACGTAG  GAGAGCGCTC  TTCGGTCTCT  AGTTTGGGGT
100   1001  TTGGAGATAG  TCGGTCAGGG  ATCTTGATGG  ACCCAAAACG  GGTGCATGCT
105   1051  CCCTTATCTG  AAGGGCATTG  TTGTCATTCC  TACCTTGAG  ACTTAGAAAA
40   1101  TCCCGGGTTA  CAAAAACAA  TTTTAGCGGC  ATTTCTGAAT  CCTAAGGAGT
     1151  TGAGCAGTAC  CATACTGCAA  CCTATATCTC  TAAATCTTAT  CTTAAATAGC
     1201  AAAACTTACT  TAAGGCAGCA  CTTTGGCTTT  TTTGAGAGGA  TGAGCAGAAG
     1251  TGATCGCAAT  GTGGTTGTCG  TTGTATGTGA  TTCTTGGTGG  GGTACCGACT
45   1301  GGAAGGAGGA  GCCAAGCTTC  CAACACTTTA  TTATGGAGCT  AGAGTGTCTG
     1351  GGGTATTTCG  ACTTCAATAT  TTTTGCCTTT  AGATCTAATA  GCATGTGTGT
     1401  AGAAGAACGT  AGGATCTTAA  ATGAAAGTTC  TCAAGAGAAA  GCCTTTACCA
     1451  TGATTTTCTG  TGAGGATTCA  GTATCTCAAG  GAGATATCCG  CTGTTTGCAT
     1501  TTGGCGTCTG  AAGGAATGCT  TTGTGGTAAA  GAGTGCTATG  CTGTCGATGT
50   1551  CTATACGTCA  GGATGCGCGA  ACTTTATGAT  GGAAGAAGTC  TTAACCTTGG
     1601  AGCGAGAATC  TAATCTGTGG  AATAGAAAGC  ATGGTCTTTG  GAAAAGAGAA
     1651  GTTAGAAAAC  AGAAACAAGA  AGCTGCTTTG  GATCAAGACG  AGAGCGAGAT
     1701  TTACGTTTGT  AATCAGCTGA  CGCGCAACA  GAACCTCGCT  TGTTCTTGA

```

The PSORT algorithm predicts inner membrane (0.2869).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 125A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 125B) and for FACS analysis.

751 GATAAGTTAG CGAGCCAAAT AGAGTTCTTA TGCCCCAGTG ACGTGTGTC  
 801 CATTTCTGGT AAAGACCCTT TGATTCTGA TGATGAGGAT GAGGAACTGA  
 851 ATCCTAAAGT TTCATCTGCT GCAGACTCTA AAGATAAAAC CTAG

The PSORT algorithm predicts cytoplasm (0.1587).

- 5 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 123A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 123B) and for FACS analysis.

These experiments show that cp6738 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## 10 Example 124

The following *C.pneumoniae* protein (PID 4376739) was expressed <SEQ ID 247; cp6739>:

1 MTHCLHGWFV VVRHHFVQAF NFSRPLYSRI THFALGVIA IPIVGHVLMG  
 51 VDWLISHCFE RGVSHPGFPS DIAPILKVEK IAGRDHISRI ENQLKSLRKT  
 101 IEVEDLDKVH GQYQENPYAD MASSEVLKLD KGVHVSELGK AFSRVRNRIT  
 15 151 RSYSYAPTPQ LDSIAIVGID LVSPEEQENL VRLANEVIQL YPKSKTTLYL  
 201 LIDFNKEWVG DISSDKEKQL RSLGLHSEVQ CLSVLEPQGA EGEDTKHFDL  
 251 MVGCGYKDSY LREGKILQQA LGTSLGTVPW VNVMHTLPSR YRSRLSLPIN  
 301 TEKDKTELYK EISRTHHQLH TLGMGLGAQD SGLLLDRQRL HAPLSQGS HC  
 351 HSYLADLTHE ELKILLFSAF VDAKNISKKE LREVSILNFAN DTSVECGCAF  
 401 YF\*

The cp6739 nucleotide sequence <SEQ ID 248> is:

1 ATGACTCATT GCTTACATGG TTGGTTTCT GTAGTTCGTC ATCACTTTGT  
 51 GCAGGCGTTT AATTTCTCAC GTCCTTTATA TTCTCGAATT ACCCACTTCG  
 101 CTTTAGGGGT GATTAAGGCC ATCCCATTTG TAGGGCATCT TGTATTGGGA  
 25 151 GTCGATTGGT TGATCTCTCA TTGCTTCGAG AGGGGAGTCT CACACCCTGG  
 201 GTTCCCTTCA GATATTGCTC CTATACTGAA AGTAGAAAAG ATCGCGGGCC  
 251 GAGATCATAT TTCTAGAATC GAAAATCAGC TAAAGAGCCT TAGGAAAAC T  
 301 ATCGAGGTTG AAGATCTAGA TAAAGTCCAC GGGCAATATC AAGAGAATCC  
 351 TTATGCAGAT ATGGCCTCTA GTGAGGTTCT TAACTCGAT AAGGGAGTTC  
 30 401 ATGTTAGCGA GCTTGGCAAA GCCTTTTCTA GAGTTCGCAA TCGCATCACC  
 451 AGATCTTATA GTTATGCCCC TACTCCTCAG TTGGACTCTA TAGCTATTGT  
 501 TGGTATAGAT CTCGTCAGTC CTGAAGAACA AGAGAATTTA GTACGCTTGG  
 551 CGAATGAGGT CATTCAACTC TATCCCAAAT CAAAGACAAC TCTATATCTT  
 601 CTTATCGATT TTAATAAGGA GTGGGTAGGG GATATCTCCT CTGATAAGGA  
 35 651 AAAACAGCTC CGTTCTCTAG GTCTACATTC TGAAGTTCAG TGTCTTTCCG  
 701 TCTTGGAACC TCAGGGTGCC GAGGGCGAAG ATACGAAACA CTTTGACCTT  
 751 ATGGTCGGCT GTTATGGGAA GGATTCCTTAC TTAAGGGAGG GTAAAATTTT  
 801 ACAGCAGGCC CTAGGGACTT CGTTAGGTAC TGTTCCTTGG GTGAATGTTA  
 851 TGCACACATT GCCATCTAGG TATAGATCTC GGCTTTCTCT ACCTATAAAT  
 40 901 ACCGAAAAGG ATAAGACAGA GCTTTATAAA GAGATTTCTC GTACACACCA  
 951 TCAGTTGCAT ACTTTGGGAA TGGGACTTGG AGCCCAGGAT TCAGGATTGC  
 1001 TCTTAGACCG GCAACGACTC CATGCTCCTT TATCTCAAGG GTCTCACTGC  
 1051 CATTCCTATC TTGCAGATCT CACCCATGAA GAGCTGAAAA TTTTGTTATT  
 1101 TTCAGCATTT GTGGATGCTA AGAACATAAG TAAGAAAGAG CTTCTGTGAGG  
 45 1151 TATCTCTAAA TTTTGCTAAC GATACTTCCG TAGAGTGTGG CTGCGCTTTT  
 1201 TACTTTTAG

The PSORT algorithm predicts inner membrane (0.2190).

- 50 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 124A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 124B) and for FACS analysis.

The PSORT algorithm predicts inner membrane (0.2338).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 126A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 126B) and for FACS analysis.

- 5 These experiments show that cp6742 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 127

The following *C.pneumoniae* protein (PID 4376744) was expressed <SEQ ID 253; cp6744>:

```

10      1  VIQHLLNFAL EETPSISVQY QEQEKLSPCD HSPEIGKKKR WNKLESFSTY
      51  CSLFMSVKDH YKLNLIQNS LSGWLLDPYR VCAPLSSPYS CPSYLLDLQN
     101  KELRRSLLST FLDPKNLTSE TFRSVSINFG NSSFGQRWSE FLSRVLHDEK
     151  EKHVAVVCND AKLLEEGLSP EALSLLLEEDL RESGYSYLN I LSVSPEGVSK
     201  VQERQILRRD LQGRSFTVMI TDLPLGSEDI RSLQLASDRI LVSSSLDAAD
     251  ACASGCKVLV YENPNASWAQ ELENFYKQVE RRR*

```

- 15 The cp6744 nucleotide sequence <SEQ ID 254> is:

```

      1  GTGATACAAC ATCTTCTAAA CTTTGCTCTA GAAGAGACCC CTTCCATTTC
      51  CGTGCAATAC CAAGAACAAG AGAAGCTCTC TCCGTGCGAT CATTCCCCAG
     101  AAATAGGTAA AAAGAAAAGA TGGAATAAGC TGGAATCCTT CTCCACGTAT
     151  TGTTCCTCTGT TTATGTCTGT TAAGGATCAT TATAAGCTGA ATCTAGGAAT
     20  201  TCAGAATTCC CTGTCAAGGT GGCTTCTGGA TCCCTATAGG GTTTGCGCGC
     251  CTTTATCTTC ACCGTACTCG TGTCTTCCT ATCTTTTAGA TTTGCAAAAC
     301  AAAGAGCTAC GTCGTTCCCT TCTGTCAACG TTTCTAGACC CTAAAAATCT
     351  CACTAGCGAA ACATTCCGTT CTGTCTCTAT AAACCTTGGC AACTCTTCGT
     401  TTGGACAGAG ATGGTCAGAG TTTCTATCTC GTGTTCTGCA CGACGAGAAA
     25  451  GAAAAGCAGC TAGCTGTTGT TTGTAATGAT GCAAACTTC TGGAAGAAGG
     501  ATTGTCCCCA GAGGCATTGT CTCTATTAGA AGAAGACTTA AGAGAATCAG
     551  GGTATTGCGT TCTAAACATT CTCTCGGTGA GCCCCGAAGG AGTCTCCAAG
     601  GTTCAGGAAC GTCAGATTCT AAGGCGAGAT CTCCAAGGAC GGTCTTTTAC
     651  TGTGATGATT ACAGATCTTC CTTTAGGTAG CGAAGATATC CGTAGTTTAC
     30  701  AATTAGCCTC GGATAGGATT TTAGTCTCCA GTTCTCTTGA TGCCGCGGAT
     751  GCATGTGCTT CGGGATGTAA AGTCTTAGTC TACGAAAATC CAAATGCATC
     801  CTGGGCTCAG GAATTGAGA ACTTCTACAA ACAAGTTGAG AGAAGAAGGT
     851  AG

```

The PSORT algorithm predicts cytoplasm (0.3833).

- 35 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 127A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 127B) and for FACS analysis.

These experiments show that cp6744 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### 40 Example 128

The following *C.pneumoniae* protein (PID 4376745) was expressed <SEQ ID 255; cp6745>:

```

      1  VACPSISSWF TVVRQHFNVA FDFTHPVCSR ITNFALGIK AIPVLGHIVM
      51  GIEWLISWIP RHTVRHGMFT SDVSSAIKVE QTRGHNC LAP LEAYLSSLRV
     101  PISQEDLGKV HGRTPEDPFV DITPTEIVQL LPDEELSTVD EALQGVRSRL
     151  TYAYRSVEKP MIQDLALVGF GLRDSADLIN FVRLANGVQN HYPHTKVKLY
     201  LAKNLADVWD CEISEEKGQ LRALGLDPKI ESISLTSAGL PSVPEVATVD
     251  FMITCYGKDQ EVQDP*

```

These experiments show that cp6741 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 126

The following *C.pneumoniae* protein (PID 4376742) was expressed <SEQ ID 251; cp6742>:

```

5      1 LFVSNFIFV VMPIFYISSW ISTVRQHFVK AFDFSRPFCS RVTNFALGVI
      51 KAIPVGHIV MGMEWLSSC VAGIITRSSF TSDVVQIVKT EKALGRDHIS
     101 RVAEILQRE GTITPENQDK VHGFVPCPF GRKSEETLK LKPGEREGTL
     151 DTVFSPIRTR VTRAYLQAPR PEIRTISIVG SKLKPQDFS QFVSLANETQ
     201 RLHPEALVCL YLTGLNRESQ MCDTPTAEKK QYLHNSGLDS RIQCKDSKED
    10  251 DAGSPENPEL WIGYYSREQQ HNIDGQYIQQ CLGKSADPIP WIHVTEDTKD
     301 FYYPPNFTSY SHTRQSTDPT SPPRLPESEG DKDSL YGQLS RSYHHEYMLG
     351 LGLKPEDAGL LMDPDRIYAP LSQGHYCHSY LADIENEDLR TLVLSPLFLDP
     401 GNLSSDLRLP VAFNIARLPL ELDSLFFRLV AGQQEGRNIV TLAHGTPRPE
     451 DLDPDMSNIL TRRLQMSGYS YLNIFYKSR KMIVKERQFF GDRSEGSFT
    15  501 LILFEDPISA ADFRCLQLAA EGMVAKDLPS VADICASGCS CIQFSEMQSP
     551 QAIEYRQWEA RVEDEAGEEA REPVIYSQDQ LSSMLTTQQN FVFSLDVAVK
     601 QAIWRFRSKG LLTMERKALG EEFLTAIFSY LGSQERNENM GKRTTEEHEV
     651 VISFEELDRM VQVLP AEVPA DSGNDPTRPV PNPDSNPDSS QNEGS*
  
```

The cp6742 nucleotide sequence <SEQ ID 252> is:

```

20      1 TTGTTTGTTT CTAATTTTAT TTTTTTTGTT GTTATGCCAA TTCCCTATAT
      51 TTCTTCTTGG ATTTCTACCG TTCGACAGCA TTTTGTTAAG GCGTTTGATT
     101 TCTCTCGTCC CTTTTGTCTT AGGGTTACGA ATTTTGCTTT AGGGGTCATC
     151 AAGGCCATCC CTATTGTAGG ACATATTGTC ATGGGGATGG AGTGGTTAGT
     201 TTCTTCTTGT GTTGCCGGA TTATTACTAG GTCCTCCTTT ACCTCAGATG
    25  251 TCGTTCAGAT TGTAAGACT GAGAAGCGT TAGGTCGAGA TCATATATCT
     301 CGAGTGGCGG AGATATTGCA AAGAGAAAGG GGGACCATAA CTCCTGAGAA
     351 TCAAGATAAG GTGCATGGA AGTTTCTTGT CTGTCTTTT GGTCTGTTAA
     401 AATCCGAGGA AACTTTAAAA CTTAAGCCGG GAGAAAGAGA GGAACCTTTA
     451 GATACTGTAT TTTCTCCGAT TCGCACGCGC GTGACTCGTG CGTACTTACA
    30  501 GGCCCCCGA CCCGAAATAC GTACGATTTC TATTGTGGGT TCGAAACTTA
     551 AAACCTCTCA AGATTTCTCG CAATTTGTGA GTCTCGCGAA TGAAACGCAG
     601 AGACTGCATC CTGAAGCGTT AGTTTGTCTG TATTTGACAG GCTTGAATCG
     651 CGAATCTCAG ATGTGCGATA CAACTACTGC AGAGAAGAAG CAGTACCTAC
     701 ATAACCTCAG TCTCGACTCT AGAATCCAGT GCAAAGACAG TAAAGAAGAC
    35  751 GACGCTGGCT CTCCTGAAAA TCCCGAACTT TGGATTGGCT ATTATTCACG
     801 AGAGCAACAG CATAATATAG ACGGGCAGTA TATTCAGCAG TGTCTAGGGA
     851 AGAGTGCAGA TCCAATTCCT TGGATTATG TTAAGTGAAG CACAAAGGAT
     901 TTTTATTACC CACCAAACCT TACTTCATAC TCACATACAA GACAATCTAC
     951 AGACCAACA TCGCCACCAA GACTCCCTGA AAGTGAGGGG GATAAGGATT
    40 1001 CCTTGTACGG ACAACTGAGT CGATCGTATC ACCATGAGTA TATGCTTGCT
    1051 TTGGGATTAA AACCAGAGGA TGCAGGACTC CTGATGGACC CGGATAGAAT
    1101 CTATGCTCCT CTATCCCAAG GGCATTATTG TCATTCTTAC CTTGCGGATA
    1151 TAGAAAAATGA GGATCTACGA ACTTTAGTCC TTTTCGCTTT CCTAGATCCT
    1201 GGCAATCTTA GTAGCGAGGA TCTTCGTCTT GTAGCATTC AATATCGTAG
    45 1251 ATTGCCATTA GAATTGGACT CGTTATTTT CCGCCTTGTT GCGGGTCAGC
    1301 AAGAAGGGAG AAACATAGTT ACCCTTGCCC ACGGAACCTC TCGTCCAGAA
    1351 GATCTTGATC CTGACTCAAT GAACATTCTG ACCAGAAGAT TACAAATGTC
    1401 TGGATATAGC TATTTGAACA TTTTCTCCTA TAAATCACGG AAAATGATTG
    1451 TAAAAGAACG TCAGTTCCTT GGAGATCGTT CTGAAGGGAA GTCTTTCACA
    50 1501 TTGATCTTAT TTGAGGATCC CATTAGTGCA GCAGATTTC GTTGTTTGCA
    1551 GCTAGCTGCA GAAGGTATGG TTGCTAAGGA TCTCCCCAGC GTAGCAGATA
    1601 TTTGTGCCTC TGGATGTTCC TGCATTCACT TTTCTGAGAT GCAGAGTCCT
    1651 CAGGCTATTG AATATAGACA ATGGGAGGCA CGTGTGCAAG ATGAAGCAGG
    1701 AGAAGAAGCC AGAGAACCAG TAATTTATTC TCAGGATCAA TTGAGCAGCA
    55 1751 TGCTCACTAC ACAACAGAA TTTGTATTTT CTCTAGATGC TGTGGTAAAA
    1801 CAGGCGATCT GGAGATTCCG TTCGAAAGGT CTTCTTACTA TGGAAAGAAA
    1851 GGCATAGGCT GAGGAGTTCT TAAGTCCGAT ATTTTCTTAT TTAGGAGTCT
    1901 AGGAGCGTAA TGAGAATATG GGGAAAAGAA CTACCGAAGA ACATGAGGTC
    1951 GTTATCAGCT TCGAAGAGCT AGATCGCATG GTGCAAGTCC TCCCAGCCGA
    60 2001 AGTCCCTGCA GATTCAGGCA ATGATCCTAC GCGTCCCGTT CCTAATCCAG
    2051 ATAGTAACCC TGATTCTCTG CAAAATGAAG GCAGTTAG
  
```

1101 TGATGAAGAT GTTCCCTCTA CCTCTGAGGA TCCTTCAGAT GATCATCCTT  
 1151 CGGATCTTGA AGACTCTTAA

The PSORT algorithm predicts inner membrane (0.1447).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 129A) and also as  
 5 a his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used  
 in a Western blot (Figure 129B) and for FACS analysis.

These experiments show that cp6747 is a surface-exposed and immunoaccessible protein, and that it  
 is a useful immunogen. These properties are not evident from the sequence alone.

### Example 130

10 The following *C.pneumoniae* protein (PID 4376756) was expressed <SEQ ID 259; cp6756>:

```

1  MASGIGGSSG LGKIPPKDNG DRSRSPSPKG ELGSHEISLP PQEHGEEGAS
51  GSSHIHSSSS FLPEQESQS SSSAASSPGF FSRVRSVDR ALKSFGNFFS
101 AESTSQARET RQAFVRLSKT ITADERRDVD SSSAAATEAR VAEDASVSGE
151 NPSQGVPEFS SGPEPQRLFS LPSVKKQSG LRLVQTVRDR IVLPSGAPPT
15  201 DSEPLSLYEL NLRSLRLRQE LSDIQSNDQL TPEEKAEATV TIQQLIQITE
251 FQCGYMEATQ SSVSLAEARF KGVETSDEIN SLCELTDPD LQELMSDGDS
301 LQNLDETAD DLEAALSHTR LSFSLDDNPT PIDNNPTLIS QEEPIYEEIG
351 GAADPQRTRE NWSTRLWNQI REALVSLGGM ILSILGSILH RLRIARHAAA
401 EAVGRCCCTCR GEECTSSEED SMSVSGSPSEI DETERTGSPH DVPRRNGSPR
20  451 EDSPLMNALV GWAHKGHAKT KESSESSTPE ISISAPIVRG WSQDSSVSFI
501 VMEDDHIFYD VPRRKDGIYD VPSSPRWSPA RELEEDVFGD YEVPITSAPD
551 SKDKNIYMTF RLATPAIYDL PSRPGSSGSS RSPSSDRVRS SSPNRGVPPL
601 PPVPSPAMSE EGSYEDMSG ASGAGESDYE DMSRSPSPRG DLDEPIYANT
651 PEDNPFQTQRN IDRILQERSG GASASFVEPI YDEIPWIHGR PPATLPRPEN
25  701 TLTNVSRLVS PGFGPEVRAA LLESVSAMV VEAESIVPPT EPGDGESEYL
751 EPLGLLVATT KILLQKGWPR GESNA*
```

The cp6756 nucleotide sequence <SEQ ID 260> is:

```

1  ATGGCATCAG GAATCGGAGG ATCTAGTGGA TTAGGAAAGA TTCCACCTAA
51  AGATAATGGG GATAGAAGTC GATCGCCCTC TCCTAAGGGA GAACCTGGCA
30  101 GCCACGAGAT TTCCCTGCCT CCTCAAGAAC ATGGAGAGGA AGGAGCTTCA
151 GGATCTTCGC ATATACATAG CAGTTCCTCT TTTCTACCAG AAGATCAGGA
201 GTCTCAGAGC TCTTCTTCGG CAGCTTCTAG CCCGGGATTT TTTTCTCGCG
251 TACGTTCTGG GGTAGACAGG GCCTTAAAAT CATTTGGCAA CTTTTTTTCC
301 GCAGAGTCTA CGAGTCAAGC GCGTGAAACG CGACAAGCTT TTGTTAGATT
35  351 ATCAAAAACC ATCACCCTGG ATGAGAGACG GGATGTCGAT TCATCAAGTG
401 CTGCTGCTAC AGAAGCCCGA GTGGCAGAGG ACGCGAGTGT TTCAGGCGAA
451 AATCTTCTC AGGGGGTTCC AGAAACCTCT TCTGGACCAG AACCTCAGCG
501 TTTATTTTCT CTTCTTCTAG TAAAAAACA GAGCGGTTTG GGTCTGGTTG
551 TACAGACAGT TCGCGATCGC ATAGTACTTC CTAGTGGGGC TCCACCTACA
40  601 GACAGCGAGC CTTTAAGTCT CTACGAGCTA AACCTCCGTT TGAGTAGTTT
651 ACGTCAGGAG CTCTCTGACA TACAAAGTAA TGATCAGTTG ACTCCAGAGG
701 AAAAAGCAGA AGCCACAGTT ACCATACAA ACCTGATCCA AATTACAGAA
751 TTCCAATGCG GCTATATGGA GGCAACACAA TCTTCGGTAT CTCTAGCAGA
801 AGCTCGTTTT AAGGGGGTAG AAAGTAGTGA TGAGATCAAT TCCCTCTGTT
45  851 CAGAACTGAC AGATCCTGAG CTTCAAGAAC TCATGAGTGA TGGAGACTCT
901 CTTCAAAACC TATTAGATGA GACTGCCGAC GATTTAGAAG CTGCTTTGTC
951 CCATACTCGA TTGAGTTTTT CTTTAGACGA TAATCCAACT CCGATAGACA
1001 ATAATCCAAC TCTGATTCTT CAAGAAGAGC CTATTTATGA GGAAATCGGA
1051 GGAGCTGCAG ATCCTCAAAG AACTCGGGAA AACTGGTCTA CAAGATTATG
50  1101 GAATCAGATT CGCGAGGCTC TGGTTTCTCT TTTAGGAATG ATTTTAAGCA
1151 TTCTAGGGTC CATCTTGAC AGGTTGCGTA TTGCTCGTCA TGCAGCTGCT
1201 GAAGCAGTGG GTCGTTGTTG CACGTGCCGA GGAGAAGAGT GTACTTCTTC
1251 TGAAGAGGAC TCGATGTCGG TGGGGTCTCC TTCAGAAATT GATGAACTG
1301 AAAGAACGGG CTCTCCGCAT GACGTTCCAC GCAGAAATGG AAGTCCACGT
55  1351 GAAGATTCTC CATTGATGAA TGCCTTAGTA GGATGGGCAC ATAAGCACGG
1401 TGCTAAAACC AAGGAGAGTT CAGAATCAAG TACCCCGGAA ATTTTCGATT
1451 CTGCTCCCAT AGTGAGAGGT TGGAGTCAAG ACAGTCCGT CAGTTTTATT
```



The cp6745 nucleotide sequence <SEQ ID 256> is:

```

      1  GTGGCTTGTC CAAGTATTTC TTCTTGGTTT ACTGTCGTTT GACAGCATTT
     51  TGTAAACGCC TTTGATTTCA CCCATCCCGT TTGTTCTCGG ATTACAAATT
    101  TTGCTTTGGG GATCATTAAAG GCAATTCCTG TATTAGGACA CATTGTCATG
5     151  GGAATCGAGT GGTGATTTC CTGGATTCCC AGACACACCG TTCGTCATGG
    201  AATGTTTACT TCTGATGTCT CTAGTGCTAT TAAAGTAGAA CAAACACGGG
    251  GTCATAATTG TTTAGCTCCC CTAGAAGCCT ATTTAAGTAG CTTGAGAGTC
    301  CCCATTTCCT AAGAAGATCT AGGCAAAGTA CACGGGAGAA CCCCAGAAGA
    351  TCCCTTCGTA GATATCACAC CCACAGAAAT TGTCCAACCT CTCCTCGATG
   10  401  AAGAACTCTC TACTGTAGAT GAGGCACTGC AAGGCGTTCG TAGTAGGTTA
    451  ACCTATGCCT ATAGGTCCGT AGAGAAACCT ATGATTCAAG ATCTTGCTCT
    501  TGTGGGTTTT GGTCTCCGAG ATTCTGCGGA CCTCATAAAT TTCGTGCGTC
    551  TTGCTAATGG CGTGCAGAA CACTATCCCC ATACTAAAGT GAAGCTCTAT
    601  TTAGCGAAGA ACTTGGCAGA TGTCTGGGAC TGTGAAATTT CTGAAGAGGA
   15  651  AAAAGGGCAA CTCCGAGCTC TAGGTTTAGA CCCTAAATA GAGAGTATAT
    701  CCTTACGAG TGCAGGTCTT CCTTCAGTGC CAGAAGTCGC TACTGTCTGAT
    751  TTTATGATTA CCTGTTACGG GAAAGATCAG GAAGTCCAAG ATCCCTAG

```

The PSORT algorithm predicts inner membrane (0.2253).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 128A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 128B) and for FACS analysis.

These experiments show that cp6745 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 129

The following *C.pneumoniae* protein (PID 4376747) was expressed <SEQ ID 257; cp6747>:

```

      1  MMKQGVGQDA KELYTFLSRG NEHYQPCLWF SLEELGFLF DEKMLCAPLS
     51  EDHYCHSYLV DLVDQHLKDL ILSMFLDPQN ISAGELLKVS INVGDSPSPL
    101  QQKDFLSMVL RDETGKNVVV VFKGVLSLPA TQVCKLVEEL NSKDYSYLN
   151  PSCHGDSSPQ LLFRKELEGT SGRYFTVICA LYLGDFTDMRS LQLASERIMV
   201  SREFDLVDAY AARCKLLKID HTNWRPGTFS RHADFADAVID VSAGFNSREF
   251  KLITQANQGI LESGELPLPS KTFWEGFLAF CDRVTVTRHF IPMLDAAIKQ
   301  AVWTHKHPSL IDKBEALDL KTQCLPSIVS YLEYVTNSHE KTSKGPFIQK
   351  EIIADCSPLK EALFPGSDED VPSTSEDPSS DHPSDLEDS*

```

The cp6747 nucleotide sequence <SEQ ID 258> is:

```

   35  1  ATGATGAAAC AAGGAGTCGG GCAGGATGCT AAAGAGCTAT ACACATTTCT
     51  ATCTCGTGGG AATGAGCATT ACCAACCCTG TCTATGGTTC AGTCTCGAAG
    101  AGGAACTCGG ATTCTTTTTC GATGAAAAAA TGCTCTGCGC CCGCTCTATCT
    151  GAGGATCACT ATTGCCACTC GTATCTTGTA GATCTAGTGG ATCAACATTT
    201  AAAGGATTTA ATATTATCGA TGTTTTTTAGA TCCTCAGAAT ATCTCAGCAG
   40  251  GAGAACTCCT CAAGGTCTCT ATAAACGTTG GAGATTCTTT TTCTCCTCTA
    301  CAACAGAAAG ATTTCTCTCT GATGGTCTTA CGTGATGAAA CGGGAAAAAA
    351  CGTCGTCTGT GTTTTTAAAG GAGTTCTCTC CTTACCCGCA ACCCAAGTCT
    401  GCAAATTAGT AGAGGAATTG AACTCTAAGG ACTACTCCTA CCTCAATATA
    451  TTTTCTTGTC ACGGAGATAG TAGTCCTCAG CTTTATTATCC GTAAGGAATT
   45  501  AGAGGGAAC T CAGGGCGTT ATTTTACAGT GATTTGCGCT TTATATCTAG
    551  GGGATACAGA CATGCGTAGT TTACAACCTG CTTCTGAAAG GATCATGGTC
    601  TCTAGAGAGT TTGATCTTGT AGATGCCTAT GCTGCAAGAT GCAAGCTCTT
    651  GAAAATCGAT CATACAAATT GGAGACCTGG AACTTTCAGT CGCCACGCCG
    701  ATTTGCGAGA TGCTGTAGAC GTATCAGCAG GATTTAAGTC AAGAGAATTT
   50  751  AAAGTATGTA CGCAGGCGAA TCAAGGGATC CTAGAGTCTG GAGAACTCCC
    801  GCTCCCTTCA AAAACCTTCT GGGAAAGGAT CTTAGCATTC TGTGATCGAG
    851  TGACTGTAC GAGACACTTC ATTTCAATGT TAGACGCCG TATAAAGCAA
    901  GCGGTATGGA CTCATAAACA TCCAGCTTG ATAGATAAAG AGTGTGAAGC
    951  CCTAGACTTG AAAACACAGT GCTTGCCATC TATCGTATCG TACCTTGAAT
   55 1001  ATGTCACAAA CTCTCACGAA AAAACATCGA AAGGCCCGTT CATACAAAAA
    1051  GAGATTATCG CAGACTGTTC TCCTCTTAAA GAGGCGCTCT TCCAGGTTTC

```

-156-

```

1101 TACAGAGGTG CTTGTTGAGA AAGTAACGGG GCAGGTTGCT ACGGGTCACT
1151 CTCCTTATTT TGAAAAGGTT TCTTTCCCTG TTGTAGGAAC GGTAGCTATC
1201 AACACTCTAG TTTCTGTGCG TCTTGATAGG GTAGAGGAAG AAGGGCTGAT
1251 TGGGGAGATT GTATGA

```

5 The PSORT algorithm predicts inner membrane (0.1574).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 131A) and also as a his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 131B) and for FACS analysis.

10 These experiments show that cp6761 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 132

The following *C.pneumoniae* protein (PID 4376766) was expressed <SEQ ID 263; cp6766>:

```

15 1 MATSVFVTSS TSVGEANSSN ERFTESTRM YYAALVLGAL SCLIFIAMIV
51 IFPQVGLWAV VLGFGALGCLL LSLAIVFAVS GLVLGKTLLEP SREATPPEIV
101 AQKEWTTQQD VLGNEYWRSE LISLFLRGDL HESLIVDSKD RSLDIDQSLQ
151 NILKLEPLST TSLLLKKDCV HINIILHLVR QWNLLGVDLS PEVTAHAEEEL
201 LLFLIEEQYY SPDILKLIRY GDALQATSPL MDWADSGSFS VDADGVFSCR
251 REECSPEDAL AQFDLLLEALE NPDRRFLKDS FLTYYIWSSSF FEKFLHRHLE
301 SLQRKLPETA IDVARYEAQI QTFLSRYFQK LDLINAMSLD WGYNCAEGEK
20 351 CYESANQRLD NLFIAFSSSV PAMKRLFDKY GSVVRVDRRQ IREQILSNTE
401 ILENESGFLC SLYEYPLSYL IDWAVLLDCV RGTBISLEDQ ADYTVCLQGL
451 DSMLSQFASR LQSGQKVLNP RDVLSEQAAV MLVHGLAAQG VSFQGLKALM
501 YLTAVPQRMW LGALPLFESF PVFNRMKEFL GESLGD*

```

The cp6766 nucleotide sequence <SEQ ID 264> is:

```

25 1 ATGGCAACCT CTGTTCTGTG AACTTCATCT ACTTCTGTAG GAGAGGCTAA
51 CTCTTCCAAC GAAAGATTTA CTGAACGAAC ATCGCGAATG TATTACGCAG
101 CTTTAGTCCT AGGGGCTTTG AGCTGTTTAA TTTTATTGTC TATGATTGTC
151 ATTTTCCAC AGGTCGGATT GTGGGCTGTG GTCTCTGGT TTGCTCTTGG
30 201 ATGTTACTT TTAAGCTTAG CTATCGTTTT TGCTGTCTCC GGTCTCGTTT
251 TAGGCAAGAC TTAGAACCT AGTCGAGAAG CGACTCCTCC AGAAATTGTT
301 GCGCAAAAGG AGTGGACTAC ACAACAAGAT GTCTTAGGGA ATGAGTATTG
351 GCGTTCCGAG TTGATTTCCT TGTCTTACG AGGGGATCTC CACGAATCTC
401 TGATTGTTGA TTCTAAGGAT CGATCTTTAG ATATTGATCA GAGTTTACAA
451 AATATATTGA AACTTGAGCC CCTATCTACG AACTTTTCGC TGTTAAAGAA
35 501 AGATTGTGTC CACATCAATA TCATTTTACA TTAGTGAGA CAGTGGAAC
551 TACTGGGAGT GGATCTTAGT CCTGAAGTCA CTGCGCACGC CGAGGAACTT
601 CTACTCTTTT TGATAGAAGA GCAGTATTAC TCTCCTGATA TTTTGAAATT
651 GATTCGCTAC GGAGATGCTT TACAAGCAAC GTCTCCTTTG ATGGATTGGG
40 701 CAGATTACAG TTCTTTTAGT GTAGACGCAG ACGGGGTATT TAGCTGTCTC
751 AGAGAAGAAT GTTCTCCTGA GGATGCTTTG GCGCAATTCT ATCTTCTTTT
801 GCGGTTGGAA AATCCCGACA GACGCTTCTT AAAGGATTCT TTTCTTACCT
851 ACATTGGGTC GTCTTCATTT TTTGAGAAGT TTTTACATCG CCATCTAGAG
901 AGCTTGCAAA GAAAGCTCCC AGAGACAGCG ATCGATGTCT CCCGCTATGA
951 AGCACAATA CAACATTTT TCTCTCGCTA TTTTCAGAAG CTCGATTGTA
45 1001 TAAACGCAAT GTCTTTAGAT TGGGGATATA ACTGTGCTGA GGGAGAAAAA
1051 TGTATTGAGA GCGCAATCA AAGATTAGAC AACCTATTTA TTGCTTTTTC
1101 TTCTTCTGTT CCTGCTATGA AGCGGCTCTT TGACAAATAT GGTCTGTGTT
1151 TACGGGTAGA TCGTAGGCAG ATTCGTGAGC AGATTCTTTC GAACACTGAA
1201 ATCTTAGAAA ATGAGTCAGG GTTCCTCTGC AGTTGTATG AATATCCTTT
50 1251 ATCCTATTG ATAGATTGGG CTGTTTGTCT AGACTGTGTT CGCGGTACCG
1301 AAATCTCTCT AGAAGATCAG GCCGATTACA CCGTTTGTCT GCAAGGCTTG
1351 GATTCTATGT TATCTCAATT TGCGAGTCGT TTACAGTCTG GACAAAAAGT
1401 ATTGAATCCT AGAGATGTTT TAAGTGAACA GGCTGCGGTT ATGCTTGTTC
1451 ATGGCTTGGC AGCACAGGGC GTGTCGTTTC AAGGATTGAA AGCTTTGATG
55 1501 TATTTGACAG CCGTTCCCA AAGAATGTGG TTAGGAGCAT TGCTTTTATT
1551 TGAATCTTTT CCTGTCTTTA ATCGGATGAA AGAATTTCTT GGGGAATCTC
1601 TGGGAGACTA G

```

-155-

1501 GTTATGGAAG ATGATCATAT TTTCTATGAT GTTCCTCGTA GAAAAGATGG  
 1551 AATCTATGAC GTTCCTAGTT CCCCTAGATG GAGTCCTGCG CGAGAGTTGG  
 1601 AAGAGGATGT TTTTGGAGAT TATGAAGTTC CTATAACCTC TGCTGAACCA  
 5 1651 TCTAAAGACA AGAACATCTA CATGACACCT AGATTAGCAA CTCCTGCTAT  
 1701 CTATGATCTT CCTTCACGTC CAGGATCGTC TGGGAAGCTCA CGTTCTCCGT  
 1751 CTTCAGATCG CGTACGAAGC AGCTCACCAA ATAGACGGGG TGTGCCTCTT  
 1801 CCTCCAGTTC CTTACCTGTC TATGAGTGAG GAGGGGAGCA TTTATGAGGA  
 1851 TATGAGCGGT GCTTCAGGTG CAGGTGAAAG TGATTATGAA GATATGAGCC  
 10 1901 GTTCCCCCTC TCCTAGAGGC GACTTGGATG AACCATATA TGCTAATACT  
 1951 CCTGAAGATA ATCCATTAC TCAGAGAAAT ATAGATAGAA TTTTACAGGA  
 2001 GAGGTCAGGC GGTGCTTCCG CTTCTCCTGT AGAGCCTATT TATGATGAGA  
 2051 TCCCATGGAT TCATGGCAGG CCCCCTGCTA CACTTCCAAG ACCCGAGAAT  
 2101 ACATTGACTA ATGTTTCGCT TAGAGTGAGC CCAGGGTTTG GACCAGAAGT  
 2151 AAGAGCCGCT TTGCTTAGCG AGAGCGTGAG TGCTGTTATG GTCGAAGCAG  
 15 2201 AGAGTATTGT TCCTCCAACA GAGCCGGGGG ACGGAGAATC AGAATATCTA  
 2251 GAGCCCTTAG GGGGACTTGT AGCTACAACG AAAATCTTAC TACAAAAGG  
 2301 ATGGCCTCGT GGAGAGTCGA ATGCTTAG

The PSORT algorithm predicts inner membrane (0.3994).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 130A). The  
 20 recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure  
 130B) and for FACS analysis.

These experiments show that cp6756 is a surface-exposed and immunoaccessible protein, and that it  
 is a useful immunogen. These properties are not evident from the sequence alone.

### Example 131

25 The following *C.pneumoniae* protein (PID 4376761) was expressed <SEQ ID 261; cp6761>:

1 MTVAEVKGTG KLVCLGCRVN QYEVQAYRDQ LTILGYQEV L DSEIPADLCI  
 51 INTCAVTASA ESSGRHAVRQ LCRQNPTAHI VVTGCLGESD KEFFASLDRQ  
 101 CTLVSNKEKS RLIEKIFSVD TTFPEFKIHS FEGKSRAFIK VQDGCNSFCS  
 151 YCIIPYLGR SVSRPAEKIL AEIAGVVDQG YREVIAGIN VGDYCDGERS  
 30 201 LASLIEQVDR IPGIERIRIS SIDPDDITED LHRAITSSRH TCPSSHLVLQ  
 251 SGSNSILKRM NRKYSRGDFL DCVEKFRASD PRYAFTTDVI VGFPGESDQD  
 301 FEDTLRIED VGFIKVHSFP FSARRRTKAY TFDNQIPNQV IYERKKYLAE  
 351 VAKRVGQKEM MKRLGETTEV LVEKVTGQVA TGHSPYFEKV SFPVVGTVAI  
 401 NTLVSVRLDR VEEGLIGEI V\*

35 The cp6761 nucleotide sequence <SEQ ID 262> is:

1 ATGACGGTTG CGGAAGTCAA AGGAACATT T AAGCTGGTCT GTTTAGGCTG  
 51 TCGGGTGAAT CAGTATGAGG TCCAAGCATA TCGCGACCAG TTGACTATCT  
 101 TAGGTTACCA AGAGGTCCTG GATTCTGAAA TCCCTGCAGA TTTATGCATA  
 151 ATCAATACGT GTGCTGTCAC AGCTTCTGCT GAGAGTTCGG GTCGTCATGC  
 40 201 TGTGCGTCAG TTATGTCGTC AGAACCTAC AGCACATATT GTTGTACACAG  
 251 GTTGTGTTGGG GGAATCTGAC AAAGAGTTT T TGCTTCTTT GGATCGGCAA  
 301 TGCACACTTG TTTCCAATAA AGAAAAATCC CGACTTATAG AAAAAATTTT  
 351 TTCTATGAT ACGACCTTCC CTGAGTTCAA GATCCATAGT TTTGAGGGAA  
 401 AGTCTCGAGC TTTTATTAAA GTTCAAGATG GCTGTAATTC TTTTTCGCTG  
 45 451 TACTGCATTA TTTCTTATT GCGGGGGCGT TCGGTTTCTC GTCCTGCTGA  
 501 GAAGATTTTA GCTGAAATCG CAGGGGTTGT AGACCAAGGA TATCGCGAAG  
 551 TTGTAATTGC AGGAATTAAT GTTGAGATT ATTGCGATGG AGAGCGTTCA  
 601 TTAGCCTCTT TGATTGAACA GGTGGACCGG ATTCCTGGAA TTGAGAGGAT  
 651 TCGAATTTCC TCTATAGATC CTGATGATAT CACTGAAGAT CTGCACCGTG  
 701 CCATCACCTC ATCGCGTCAC ACTTGTCTCT CGTCACACCT TGTCTCTCAA  
 751 TCGGGGTCGA ATTCAATTTT AAAGAGAATG AACCGGAAGT ATTCTCGCGG  
 801 AGATTTTTTA GATTGTGTAG AGAAGTTCCG TGCTTCTGAT CCTCGCTATG  
 851 CCTTTACTAC AGATGTGATT GTCGGATTTC CTGGAGAGAG TGATCAAGAT  
 901 TTTGAAGATA CTTTGAGAAT TATTGAAGAT GTAGGCTTTA TTAAAGTGCA  
 55 951 TAGTTTCCCT TTCAGTGCTC GTCGTGCTAC TAAGGCATAT ACTTTTGATA  
 1001 ATCAGATTCC CAATCAGGTG ATCTATGAGA GGAAGAAGTA TCTTGCTGAG  
 1051 GTTGCTAAGA GGGTAGGCCA GAAAGAGATG ATGAAGCGTT TAGGAGAGAG

-158-

```

1  ATGTCATCAC TACTGAGCTG CGGAAGAATA GAGCCGACTC GGGTTACCTG
51 TAGCTTAAAG ACGTATCTTG AGGATACGAG TCAGAATCAG TTGAGCACAC
101 GTCTAGTTCG GGCAAGTGTC ATCTTTTAT GCGCATGTGT GATCATTTTG
5   151 GTTTGTGTGG CCCTCTCTAG TTTGATTCCA AGCATTATGG CCTTGCGGAC
201 CTCTTTTACG GTAATGGGGT TAATTCCTTT TGTGATGTCA CTTCTTGGTG
251 ACGTTGCAAT TATAAGTTAT CTTACTTATA GCACTGTTAC GAGTTACCGG
301 CAAAATAAGA GAGCTTTTGA GATTCACAAG CCCGCTCGCT CCGTTTACTA
351 CGAGGGGGTC CGCCATTGGG ATTTAGGACG ATCATCTTTA GGCACAGGCG
10  401 AGATTCCCTAT AGTAAGGACG TTATTCTCTC CATTCAGAA CCATGGTCTT
451 AACCATGCCT TAGCTGCTAA AATTTTCCTA TTTATGGAGC ATTTTCAGCC
501 TGAGCCACCG AACGAGCCTT TGGTGGATTG GGCCTGTTTG ATTCGGGATT
551 TTAGGCCTCA CGTCAGTTCT TTGTGCTTTG TTATTGAAAA ACAAGGGTCA
601 TCGCTGAGGA CTAAGGAAGG CAATACGATT TGTGAGGCTT TCCGCTCTGA
651 TTACGACGCC CATTTTGCTA TGGTAGATTG CTACCGGTTG ATCCACTCTA
15  701 AGTTGATTAT AGAGAAAATG GGATTGAAGA ATATCGATAT CATTCGAGT
751 GTCATGGTTC GTGAAGATTA TCCTAGCCGT CCTGGGGAGG GCTATCGCGA
801 AGGCCTATTA CGTATGTATG GTGGCAAGGG GGCTCTGTGA

```

The PSORT algorithm predicts inner membrane (0.711).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 134A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 134B) and for FACS analysis.

These experiments show that cp6805 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 135

25 The following *C.pneumoniae* protein (PID 4376813) was expressed <SEQ ID 269; cp6813>:

```

1  MSGPSRTESS QVSVLSYVPR DKEIAPKKQF TIAKISTLAI LASLALGALV
51 AGISLTIVLG NPVFLALLIT TALFSVVTFL VYHQMTSKVS SNWQKVLEQN
101 FKPLGKAWQE KNVDCYSNEM QFYNNHLNPK FKVAIQTDAS QPFQPTFLTG
151 LRVIEKNQST GIIFNPVGP T NLIDNTATNL STILYSTLKD KSVWDTCKQR
30  201 EGGPAKGEDP FSPTEVRVVK LPNEALDQTF NLNLSSAEKK SILPTFLGHV
251 CGPKSEELPN QQEYYRQALL AYENCLKAAI ESHAAIVALP LFTSVYEVPP
301 EEILPKEGTF YWDNQTQAFK KRALLDAIQN TALRYPQRS LVLILQDPFNT
351 IESQSRSEE*

```

The cp6813 nucleotide sequence <SEQ ID 270> is:

```

35  1  ATGTCAGGAC CCTCACGTAC TGAGAGCTCT CAAGTTTCTG TACTATCCTA
51  TGTGCCTCGG GATAAAGAAA TTGCTCCTAA AAAACAGTTT ACCATAGCAA
101 AAATATCCAC TCTTGCAATC CTAGCTTCTT TAGCTTTAGG AGCTTTGGTG
151 GCTGGAATCT CTTTAACGAT AGTATTAGGG AACCTGTAT TTTTGGCTCT
201 TCTCATTACC ACGGCCCTCT TCTCAGTTGT AACCTTCTTA GTCTACCACC
40  251 AAATGACCTC AAAGGTATCT TCTAACTGGC AGAAAGTTCT AGAGCAAAAC
301 TTCAAGCCTT TGGGAAAAGC GTGGCAAGAA AAAAACGTAG ACTGCTACTC
351 AAACGAGATG CAATTTTACA ATAATCACCT GAACCTAAG TTCAAGGTAG
401 CGATACAAAC AGATGCGTCT CAACCATTTT AGCCTACTTT CTTAACTGGA
45  451 CTTAGAGTGA TCGAAAAAAA TCAATCCACA GGGATCATCT TTAATCCCGT
501 AGGCCCCAAG AATCTGATCG ACAACACTGC AACGAACCTC TCTACTATTCC
551 TTTACTCCAC CCTAAAAGAT AAAAGCGTGT GGGATACATG CAAGCAACGC
601 GAAGGGGGTC CCGCAAAAGG AGAAGACCCC TTTTCCCGTA CCGAAGTGAG
651 AGTAGTAAAA CTTCCAAACG AAGCTCTAGA TCAAACGTTT AATCTAAATT
701 TAAGCTCTGC AGAAAAGAAA AGTATTCTTC CGACCTTTT AGGCCACGTA
50  751 TGCGGCCCTA AATCTGAAGA GTTACCAAAT CAGCAAGAA ATTATCGCCA
801 AGCTTTACTA GCGTACGAGA CTGCCTTAA AGCAGCTATA GAAAGTCATG
851 AGCAATCGT TGCTCTTCCT CTCTTTACTT CGGTCTATGA AGTGCCTCCA
901 GAAGAGATTC TTCCTAAAGA AGGCACTTTC TATTGGGACA ACCAACTCA
951 AGCGTTTTGC AAACGCGCTT TATTGGACGC TATTCAAAAT ACGGCCCTAC
55  1001 GCTATCCTCA AAGATCTTTA CTTGTTATAC TCCAAGATCC TTTTAATACT
1051 ATAGAATCAC AAAGTCGTTT TGAGGAGTAA

```

The PSORT algorithm predicts inner membrane (0.6158).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 132A) and also as a his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 132B) and for FACS analysis.

- 5 These experiments show that cp6766 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 133

The following *C.pneumoniae* protein (PID 4376804) was expressed <SEQ ID 265; cp6804>:

```

10      1  MSNQLQPCIS  LGCVSYINSF  PLSLQLIKRN  DIRCVLAPPA  DLLNLLIEGK
      51  LDVALTSSLG  AISHNLGYVP  GFGIAANQRI  LSVNLYAAPT  FFNSPQPRIA
     101  ATLESRSSIG  LLKVLCHRHW  RIPTPHILRF  ITTKVLRQTP  ENYDGLLLIG
     151  DAALQHPVLP  GFVTYDLASG  WYDLTKLPFV  FALLHSTSW  KEHPLPNLAM
     201  EEALQQFESS  PEEVLKEAHQ  HTGLPPSLQ  EYALCQYRL  GEEHYESFEK
     251  FREYYGTLYQ  QARL

```

15 The cp6804 nucleotide sequence <SEQ ID 266> is:

```

      1  ATGTCTAACC  AACTCCAGCC  ATGTATAAGC  TTAGGCTGCG  TAAGTTATAT
     51  TAATTCCTTT  CCGCTGTCCC  TACAACTCAT  AAAAAGAAAC  GATATTCGCT
    101  GTGTTCTTGC  TCCCCCTGCA  GACCTCCTCA  ACTTGCTAAT  CGAAGGGAAA
    151  CTCGATGTTG  CTTTGACCTC  ATCCCTAGGA  GCTATCTCTC  ATAACCTGGG
    201  GTATGTCCCC  GGCTTTGGAA  TTGCAGCAAA  CCAACGTATC  CTCAGTGTA
    251  ACCTCTATGC  AGCTCCCACT  TTCTTTAACT  CACCGCAACC  TCGGATTGCC
    301  GCAACTTTAG  AAAGTCGCTC  CTCTATAGGA  CTCTTAAAG  TGCTTTGTGC
    351  TCATCTCTGG  CGCATCCCAA  CTCCTCATAT  CCTAAGATTC  ATAACCTACAA
    401  AAGTACTCAG  ACAAACCCCT  GAAAATTATG  ATGGCCTCCT  CCTAATCGGA
    451  GATGCAGCGC  TACAACATCC  TGTACTTCCT  GGATTTGTAA  CCTATGACCT
    501  TGCCTCGGGG  TGGTATGATC  TTACAAAGCT  ACCTTTTGTA  TTTGCTCTTC
    551  TTCTACACAG  CACCTCTTGG  AAAGAACATC  CCCTACCCAA  CCTTGCGATG
    601  GAAGAAGCCC  TCCAACAGTT  CGAATCTTCA  CCCGAAGAAG  TCCTTAAAGA
    651  AGCTCATCAA  CATAACAGTC  TGCCCCCTTC  TCTTCTTCAA  GAATACTATG
    701  CCCTATGCCA  GTACCGTCTA  GGAGAAGAAC  ACTACGAAAG  CTTTGAAAAA
    751  TTCCGGGAAT  ATTATGGAAC  CCTTACCAA  CAAGCCCGAC  TGTA

```

The PSORT algorithm predicts inner membrane (0.060).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 133A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 133B) and for FACS analysis.

These experiments show that cp6804 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 134

The following *C.pneumoniae* protein (PID 4376805) was expressed <SEQ ID 267; cp6805>:

```

40      1  MSSLLSCGRI  EPTRVTCSLK  TYLEDTSQNO  LSTRLVRASV  IFLCALLIIL
     51  VCVALSSLIP  SIMALATSFT  VMGLILFVMS  LLGDVAIISY  LTYSTVTSYR
    101  QNKRAFEIHK  PARSVYEGV  RHWDLGRSSL  GTGEIPIVRT  LFSPPQNHGL
    151  NHALAAKIFL  FMEHFSPEPP  NEPLVDWACL  IRDFRPHVSS  LCFVIEKQGS
    201  SLRTKEGNTI  CEAFRSDYDA  HFAMVDCYRL  IHSKLIIEKM  GLKNIDIIPS
    45  251  VMVREDYPSR  PEGGYREGLL  RMYGGKGAL*

```

The cp6805 nucleotide sequence <SEQ ID 268> is:

```

201 CIGFFGINGI CSTFLMLTNP RSRDRWRNL RIMVLCYRSL GSGMNLFDLS
251 NNVRMAARRH VTSCTVALYA MVTLFGWTVA IQDALQYGFP SVRDAFYRYC
301 LRHRYCLTQR NEDSLQTTGT RFQVTRTHLE DQQMVASILN LSVFGLFFGF
351 VGLMTTFGGL EISPSRWDA ANNRTVGIF*

```

5 The cp7201 nucleotide sequence <SEQ ID 274> is:

```

1   GTGCTCGTTG GTATCTGTCC TTCTCTATAT CCAGAACATC CTCGCTCCTT
51  TTATTATCGT GTTCTGGAG ATATAGGCTC CCGATTGAC GATAGAGGAT
101 TTGTAAACTC TGGAGTCGAA ACCCTGCCAT ACTCTTCAGG CAGCTTTGGG
151 ATTTTTTGGG TCTCGTTTAC GGATCCCA CA TTTAATTTTG CTATCGTAAA
201 TACCTTTATG CGAACTGCAG GGATCAATGA AGTCTCTAGA CCCATGACAC
251 AAGATACAGA AACTTCATTG ATAGAAATGA GAGACCTAAG TGAACAACAA
301 GAAGCGAATA ACACAGATTC TTTAGAGCAA GAAGAGAGCT TAATGGGTAT
351 TGTAGGACAT ACTGTGGGAG GAGTTTCCAT GACCGTGACC TCCAGTCCAA
401 ATATCTTTTA TCGTATACAA ACACCTCTGG GACTGCCAGA GACTCTTGCA
15 451 GAAGCTGAAG AAAATCCGTA CTTCCTCAAT TCTACTATAG ATAGCCTTGC
501 AGAAATAATG ATGAACCTCG TAAGGATCTC TGATGCTGTC TCTATTTTCT
551 GGATTTTTTC TATCGTAGAT ACTACATATA ATGGAGTTT ATTAGCCGTC
601 TGTATCGGCT TCTTCGGAAT CAATGGGATT TGTTCACGT TCCTTATGCT
651 TACGAATCCA CGCTCTCGTC GAGATAGATG GAGGAATTTA CGCATCATGG
20 701 TTCTTTGCTA TCGTCTTTG GGAAGCGGAA TGAATCTCTT TGATCTTAGC
751 AATAATGTGC GCATGGCAGC ACGTAGGCAT GTGACATCAT GTACAGTAGC
801 TCTCTATGCT ATGGTCACTC TATTTGGATG GACAGTAGCA ATACAAGTAG
851 CTTTGCAATA TGGTTTCCCT AGCGTTCGGG ATGCCTTCTA TAGATATTGC
901 TTACGCCACA GATATTGCTT AACTCAAAGA AACGAAGACT CTCTGCAAAAC
25 951 TACAGGAACG CGCTTTCAGG TTACCCGTAC ACATCTAGAA GATCAACAGA
1001 TGGTGGCTTC TATTTTGAAT TTGAGTGTTT TTGGGCTCTT TTTTGGATTC
1051 GTAGGGCTAA TGACCACGTT TGGAGGATTA GAAATCTCAC CATCTTGTCG
1101 GTGGGATGCA GCAAATAACC GAACGGTAGG TATTTTTTAG

```

The PSORT algorithm predicts inner membrane (0.3102).

30 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 137A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 137B) and for FACS analysis.

These experiments show that cp7201 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### 35 Example 138

The following *C.pneumoniae* protein (PID 4377251) was expressed <SEQ ID 275; cp7251>:

```

1   MAPIHGSNAF VEDILHSHPS PQATYFSSTR AQKLHEFKDR HPVLTRIASV
51  IIKIFKVLIG LIILPLGIYW LCQTLCTNSI LPSKNLLKIF KKQPNRTKLK
101 TNYLHALQDY SSKNRVASM RVPILODNVL IDTLEICLSQ APTNRWMLIS
40 151 LGSDCSLEEI ACKEIFDSWQ RFAKLIGANI LVYNYPGVMS STGSSSLKDL
201 ASAHNICTRY LKDKEQGPQA KBIITYGYSL GGLIQAEALR DQKIVANDDT
251 TWIAVKDRCP LFISPEGFHS CRRIGKLVAR LFGWGTKAVE RSQDLPCLEI
301 FLYPTDSLRR STVRQNKLLA PELTLAHAIAK NSPYVQNKEF IEVRLSSDDI
351 PIDSKTRVAL ATPILKKLS*

```

45 The cp7251 nucleotide sequence <SEQ ID 276> is:

```

1   ATGGCTCCAA TTCACGGAAG TAATGCGTTT GTTGAGGATA TTTTACATTC
51  CCACCCTTCT CCACAAGCGA CTTATTTTTC TTCAACACGC GCCCAAAAC
101 TTCATGAGTT TAAAGACAGG CATCCCGTGC TTACACGGAT TGCTTCTGTA
151 ATTATTAAAA TTTTAAAGT TCTGATAGGG CTGATCATCC TTCCCTTAGG
50 201 AATCTACTGG CTATGTCAA CGCTTTGTAC AAACCTGATT CTCCCTTCCA
251 AGAATTTATT AAAAATTTTC AAGAAGCAAC CCAACACTAA AACCTTAAAA
301 ACTAATTATT TGCATGCTTT GCAAGATTAT TCCTCGAAAA ACCGCGTTGC
351 TTCCATGAGA CGAGTTCCTA TCCTCCAGGA TAATGTTCTC ATCGACACTT
401 TGGAAATATG CCTTTCACAA GCACCTACGA ATCGTTGGAT GCTCATTTCT
55 451 TTAGGAAGTG ACTGTAGCTT GGAAGAAATC GCTTGTAAGG AGATCTTTGA

```

The PSORT algorithm predicts inner membrane (0.4291).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 135A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 135B) and for FACS analysis.

- 5 These experiments show that cp6813 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 136

The following *C.pneumoniae* protein (PID 4376844) was expressed <SEQ ID 271; cp6844>:

```

10      1  MWRVVLRLFI IFILGRAVFP LRASESFSWE TSTCLTVLGI PFIDIILTTN
      51  EDFVAQCGLQ IGTISSTNNA KIKEIFLIYK EKFPEASISF KRKEPLNLSQ
     101  SHLSDLGILC MRNGETYABG MANKENGPAI KQPKDLRLVL RCPNQPDILL
     151  YSEKEAEKGI ETNTCLCNQG YTLLDGQLIL YGDSIEKFLK BTKRKNNHTL
     201  VDLCDSDQVVT TFLGRFWSLL NYVQVFLSE DSAKILAGIP DLAQATQLLS
     251  HTVPLLFIYT NDSIHIEQG KESSFTYNQD LTEPILGFLF GYINRGSMEY
     301  CFNCAQSSLG ET*
```

The cp6844 nucleotide sequence <SEQ ID 272> is:

```

20      1  ATGTGGCGCG TTGTCCTCAG ATTCCTTATA ATTTTATCT TGGGAAGAGC
      51  CGTCTTCCCT CTAAGAGCTT CAGAAAGCTT CTCCTGGGAA ACATCGACCT
     101  GTTTAACAGT GCTAGGGATT CCTTTCATAG ATATTATCCT CACAACGAAT
     151  GAGGACTTTG TTGCCAGTG CGGCCTGCAA ATAGGAACCA TTTCTTCGAC
     201  TAATAACGCA AAAATAAAAG AAATTTT TTTT GATATATAAG GAAAAATTTT
     251  CAGAAGCCTC TATCAGTTTC AAACGAAAAG AACCTCTAAA CCTTTCCTCAA
     301  TCCCATCTCT CCGATTAGG TATTTTATGT ATGCGTAACG GAGAACTTA
     351  CGCTGAGGGA ATGGCAAATA AAGAAAACGG ACCCGCTCTA AAACAACCCA
     401  AGGATCTAAG ATTAGTTTTA CGTTGTCCTA ACCAACCAGA TACCCTGCTC
     451  TACTCGGAAA AAGAAGCAGA AAAGGGCATA GAAACAAATA CTTGCCTATG
     501  CAATCAGGGA TACACACTCC TGGATGGGCA ATTGATTCTC TACGGGGATA
     551  GTATAGAAAA GTTTCGTGAAA GAGACCAAAA GAAAGAATAA CCACACGCTT
     601  GTTGATCTTT GTGACTCACA AGTCGTGACC ACGTTCCTCG GTCGCTTTTG
     651  GTCTCTTCTA AACTACGTTT AAGTTCTTTT CCTATCTGAA GACTCCGCTA
     701  AAATTCTTGC GGGCATCCCA GACCTAGCTC AAGCTACGCA ATTGCTTTCC
     751  CACACCGTAC CTTTGCTTTT TATTTATACC AACGATTCTA TTCACATCAT
     801  AGAACAAGGC AAAGAAAGTA GTTTTACCTA TAACCAAGAT TTAACAGAGC
     851  CCATTTTAGG ATTTCTCTTT GGTTACATAA ATCGCGGCTC TATGGAATAC
     901  TGCTTTAATT GTGCACAGTC TTCATTAGGA GAAACCTAA
```

The PSORT algorithm predicts inner membrane (0.1786).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 136A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 136B) and for FACS analysis.

- 40 These experiments show that cp6844 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 137

The following *C.pneumoniae* protein (PID 4377201) was expressed <SEQ ID 273; cp7201>:

```

45      1  VLVGICPSLY PEHPRSFFYR VSGDIGSRFD DRGFVNSGVE TLPYSSGSFG
      51  IFWISFTDPT FNFAIVNTFM RTAGINEVSR PMTQDTETSL IEMRDLSEQQ
     101  EANNDSLEQ EESLMGIVGH TVGGVSMFTV SSPNIFYRIQ TLLGLPETLA
     151  EAEENPTFPN STIDSLAEIM MNLVRISDAV SIFWIFPIVD TTYNGVLLAV
```

1251 CGTCTTTAAA TCCATGCAAA AAGCAGATCC AGAAACCAAA GCTTTAATCC  
 1301 GTGAGTTTGC TCTAGATATA TTATATGCAT CCTTACGGCT TCCTCAAACCT  
 1351 TCCGCTCATA CCGAGGTCTT TTCTACACTC TTAATGGACC CAGAGACCTA  
 1401 TGAACCTAAT AAAGCTTGTA TCGCCTACTT GCTCTATGTA TTAAAGATCA  
 1451 TCGAACTATA A

The PSORT algorithm predicts inner membrane (0.5989).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 139A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 139B) and for FACS analysis.

10 These experiments show that cp7288 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 140

The following *C.pneumoniae* protein (PID 4377359) was expressed <SEQ ID 279; cp7359>:

15 1 MPGSVSSPPL SPVIVRERV SSSGSDLIQP HAVLKISILI FALVTILGIV  
 51 LVVLSSALGA LPSLVLTVSG CIAIAVGLIG LGILVTRLIL STIRKVDAMG  
 101 YDAAVKEEQY LSRIRELESE NREIRDNRRA VEDQCAHLSE ENKDLRDPEY  
 151 LHGMTERLIA SLEIENQALV AENILLKDOWN ASLSRDFRAY KQKFPLGALE  
 201 PWKEDIACIM EQNLFLKPEC IAMVKSLEPLE TQRLFLYPKG FQSLVNRFP  
 251 RSRFFQTPKY EYNSRNENED GKVAAVCARL KKEFFSAVLG ACSYEELGGI  
 20 301 CERAVALKET LPLPEAVYDT LVQEFPNLLT AESLWKEWCF YSYPYLRPYL  
 351 SVDYCKRFLV QLFEELCLKL FTTGSPEDQA LVRLFSYYRN HIPAVLASFG  
 401 LPPPETGGSV FVLLPKQENL LWSQIEVLAT RYLKDTFVRN SEWTGSFEMM  
 451 FSYNEMCKEI SEGRIRFAED YETRHSEEFPS PSLPSEEGEG BEFLPPCSEE  
 501 EVSVLERPDL DVDSMWVWHP PVPKGPL\*

25 The cp7359 nucleotide sequence <SEQ ID 280> is:

1 ATGCCAGGTT CTGTGTCATC ACCTCCTTTG TCTCCTGTAA TTGTCCGTGA  
 51 AAGGGTCCCA TCCTCTTCAG GATCCGACCT CATAACGCTT CATGCTGTTT  
 101 TAAAGATCTC CATCCTAATT TTTGCGCTTG TGACAATTTT AGGAATTGTT  
 151 CTTGTAGTGT TGTCTAGTGC TTTAGGAGCT CTTCTAGTGT TAGTTTTGAC  
 20 201 GGTTCCTGGT TGTATTGCAA TAGCTGTAGG CCTGATTGGT TTAGGGATTG  
 251 TTGTGACACG GCTGATTCTC TCTACGATCA GAAAAGTAGA TGCCATGGGT  
 301 TATGATGCTG CGGTCAAAGA AGAGCAGTAT TTGTCACGTA TCAGAGAATT  
 351 AGAGTCTGAA AATAGAGAGA TTAGAGATAG AAATCGTGCT GTCGAAGATC  
 401 AGTGTGCCCA TTTATCCGAA GAGAACAAGG ACCTTAGGGA TCCCGAATAT  
 35 451 CTACATGGAA TGAAGTAAAG GCTCATTCGG AGCTTAGAAA TAGAGAATCA  
 501 AGCTCTCGTA GCTGAGAACA TTCTTCTCAA AGACTGGAAT GCAAGCCTAT  
 551 CTAGAGATTT CCGCGCATAT AAGCAAAAAT TTCCTCTTGG GGCATTAGAA  
 601 CCCTGGAAAAG AAGATATTGC ATGTATCATG GAACAAAATC TCTTTTTTAA  
 651 ACCGGAATGT ATCGCGATGG TTAAGTCTCT TCCATTAGAG ACGCAACGGC  
 40 701 TGTTTTTATA TCCAAAAGGA TTTCAGTCTT TAGTTAATCG ATTTGCTCCG  
 751 CGGTCTCGCT TTTTCCAGAC TCCAAAAGTAT GAATATAACA GTAGGAATGA  
 801 AAATGAGGAC GGAAAGGTAG CCGCAGTGTG CGCCCGTTTG AAAAAAGAAT  
 851 TCTTCAGTGC TGTTTTAGGA GCCTGTAGTT ACGAAGAACT AGGGGGCAT  
 901 TGTGAAAGAG CAGTAGCACT TAAAGAGACG TTGCCATTGC CTGAAGCTGT  
 45 951 CTATGATACC CTAGTTCAGG AGTTCCCAA TCTTCTTACT GCTGAGAGTT  
 1001 TATGGAAAAG ATGGTGCTTC TATTCCTATC CCTACCTTCG TCCCTATCTT  
 1051 TCTGTGGATT ACTGTAAGAG GTTATTTGTA CAACTTTTTG AGGAAGTCTG  
 1101 CCTAAAGCTT TTTACAACGG GATCTCCAGA AGACCAAGCT TTGGTTCGCC  
 1151 TTTTCTCTTA CTATAGGAAT CATATCCCG CAGTCTTGGC CTCATTTGGT  
 50 1201 TTGCCCCCGC CTGAGACAGG GGGGTCTGTA TTTGTATTGC TACCAAAACA  
 1251 AGAAAACCTT CTTTGGAGTC AAATTGAGGT GCTGGCTACA AGGTATCTCA  
 1301 AAGATACCTT CGTGAGAAAC TCAGAATGGA CGGGCTCTTT CGAGATGATG  
 1351 TTTTCTTATA ACGAGATGTG TAAGGAGATC TCCGAAGGAA GGATTCTGTT  
 1401 TGCTGAAGAC TATGAAACGA GGCATTCCGA AGAATCCCTT CCTTCCCTC  
 55 1451 TCTCTGAAGA AGGAGAGGGC GAAGAATTCC TTCCTCCTTG CTCTGAAGAA  
 1501 GAGGTTTCGG TTCTTGAGCG CCCAGATCTA GATGTAGACT CTATGTGGGT  
 1551 CTGGCATCCG CCGGTCCCTA AGGGACCTCT TTAA



-161-

501 TTCTTGGCAA AGATTGCCA AGTTGATAGG GGCCAATATA CTCGTTTATA  
 551 ACTACCCCGG AGTCATGTCC AGCACAGGGA GCAGCAGCCT AAAGGACCTA  
 601 GCATCAGCTC ATAATATTTG TACAAGATAC CTTAAAGATA AAGAACAGGG  
 5 651 CCCTGGAGCA AAAGAAATCA TTACCTATGG GTACTCCCTA GGAGGTTTGA  
 701 TACAAGCAGA AGCATTGCGA GACCAGAAGA TTGTTGCAAA CGATGATACT  
 751 ACTTGGATAG CAGTCAAAGA TAGGTGTCTT CTCTTTATAT CTCCAGAAGG  
 801 TTTCCACAGT TGCAGACGCA TAGGAAAGCT AGTAGCTCGT CTTTGTGGCT  
 851 GGGGGACCAA AGCCGTAGAG AGAAGCCAAG ACCTTCCCTG CCTAGAAATT  
 901 TTTCTCTATC CTACGGATTC CTTACGAAGA TCAACAGTCA GACAGAACAA  
 10 951 GCTCTTAGCA CCTGAACTTA CTCTCGCTCA TCGGATAAAA AATAGTCCCT  
 1001 ATGTTCAAAA TAAAGAATTT ATAGAAGTAC GATTATCGTC TGATATCGAT  
 1051 CCCATCGACA GCAAACAAG AGTGGCTCTT GCCACACCAA TTTTGAAAAA  
 1101 GCTCTCTTAG

The PSORT algorithm predicts inner membrane (0.4545).

- 15 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 138A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 138B) and for FACS analysis.

These experiments show that cp7251 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## 20 Example 139

The following *C.pneumoniae* protein (PID 4377288) was expressed <SEQ ID 277; cp7288>:

1 MHMSNPISLF SPAELIAKYN LIPKTSPIYP RRELIILEE NACQTRLTNV  
 51 AQVLHPSSLF SMSKKILNPC GCSGGPLCWV ILNILAFIIT SVLFIIILLPV  
 101 NLIVAGLRLF MPLPPKKIVE DLSEPTTEET NEVIQPFIFA LQALIFEDNK  
 25 151 LRSFKIVEQS VGKAPLPNPF LNRLVAISPQ ESQEAMRKIP DLCSQLKKVL  
 201 KSLGVLTPFW KHLKYFEGFL KNEHDSNPDK KTFPILIKLL IEALTGKSSL  
 251 PKTPSTKEKM QAALFIASSC KTCKPTWGEV ITRSLNRLYS IANEGDNQLL  
 301 IWVQEFKERE LMSIQDGDDA EYRFFAAQH GERYTEAIEQ VLRNESAACL  
 351 QWHVINTMKF FHGKNLGLVT EHLQDTLGAL TLRQTTVDTH QGREDAADLSA  
 30 401 ALFLNKYLNS GNQLVNSVFK SMQKADPETK ALIREFALDI LYASLRLPQT  
 451 SAHTEVFSTL LMDPETYEPN KACIAYLLYV LKIIEL\*

The cp7288 nucleotide sequence <SEQ ID 278> is:

1 ATGCATATGT CTAACCCCAT CTCTTTGTTT TCCCCTGCAG AGTTAATAGC  
 35 51 AAAGTACAAAT TTAATTCCAA AAACCTCGCC GATTTATCCT CGGAGGACGG  
 101 AACTTATTAT CTTGGAAGAA AATGCGTGTC AAACACGCCCT AACCAACGTC  
 151 GCTCAGGTCC TACATCCTTC TAGCCTATTC AGTATGTCAA AAAAAATACT  
 201 GAATCCCTGC GGGTGCTCTG GTGGTCCCTT ATGTTGGGTG ATTCCTCAACA  
 251 TCCTAGCATT TATTATTACT TCAGTACTGT TTATCATTCT TTTACCGGTG  
 301 AATCTCATCG TAGCAGGTCT TCGTCTCTTC ATGCCTCTTC CCCCTAAAAA  
 40 351 AATCGTAGAG GATTTAAGTG AACCTACTAC TGAAGAAACG AATGAGGTCA  
 401 TTCAACCCCT CATTTCGCT TTGCAAGCGT TGCTTTTGA GGATAACAAA  
 451 CTTCGCTCTT TTAATAATGT TGAACAAAGT GTAGGCAAAG CACCCTTACC  
 501 TAATCCCTTT TTAATAAGAC TAGTAGCAAT TTCGCCGCAA GAAAGCCAAG  
 551 AAGCCATCGG GAAGATTCCG GATCTATGCT CACAACCTGAA AAAAGTATTA  
 45 601 AAGTCTCTAG GCGTGCTAAC TCCAGAATGG AAGCACATGC TGAAGTACTT  
 651 TGAGGGACTG AAAAACGAAC ATGATAGTAA TCCTGATAAA AAGACGTTC  
 701 CAATATTGAT CAAGCTCCTC ATAGAAGCTC TTACTGGAAA GTCCTCTTTA  
 751 CCCAAAACTC CTAGTACAAA GGAAAAATG CAAGCGGCTT TATTTATTGC  
 801 AAGTCTCTGC AAGACTTGTA AGCCGACTTG GGGAGAAGTC ATAACCAGAT  
 50 851 CTCTTAACAG ACTCTATAGT ATAGCTAATG AAGGAGACAA TCAGCTTCTG  
 901 ATTTGGGTTC AAGAGTTTAA AGAACGAGAG CTGATGTCCA TCCAAGATGG  
 951 TGATGATGCT GAAGAGTATC GGTTTGCGGC TCAGCAACAC GGTGAGCGTT  
 1001 ACACAGAGGC AATAGAACAA GTTCTACGAA ACGAGTCAGC AGCCAAACTA  
 1051 CAATGGCATG TGATCAACAC TATGAAATTC TTCCATGGGA AAAATCTCGG  
 55 1101 TCTAGTTACA GAACACCTAC AAGATACTCT CGGCGCCCTA ACTTTACGTC  
 1151 AAACACAGT GGACACACAT CAAGGCAGAG AAGACGCTGA TTTGTACGCT  
 1201 GCTCTTTTCC TAAATAAGTA TTTAAATCTT GGAAATCAAC TTGTTAATAG

**Example 142**

The following *C.pneumoniae* protein (PID 4377377) was expressed <SEQ ID 283; cp7377>:

```

1  MREETVSWSL EDIREIYHTP VFELIHKANA ILRSNFLHSE LQTCYLISIK
51  TGGCVEDCAY CAQSSRYHTH VTPEPMMKIV DVVERAKRAV ELGATRVCGLG
101 AAWRNAKDDR YFDRVLAMVK SITDLGAEVC CALGMLSEEQ AKKLYDAGLY
151 AYNHNLDSPP EFYETIITTR SYEDRLNLTLD VVNKSGISTC CGGIVGMGES
201 EEDRIKLLHV LATRDHIPES VPVNLLWPID GTPLQDQPPI SFWEVLRTIA
251 TARVVFPRSM VRLAAGRAFL TVEQOTLCFL AGANSIFYGD KLLTVENNDI
301 DEDAEMIKLL GLIPRPSFGI ERGNPCYANN S*

```

10 The cp7377 nucleotide sequence <SEQ ID 284> is:

```

1  ATGCGTGAAG AAACGTATATC CTGGTCATTA GAAGACATCC GCGAAATTTA
51  TCACACTCCC GTATTTGAGC TGATTCACAA AGCCAATGCC ATATTGCGTA
101 GTAATTTTCTT CCATTCAGAA CTGCAGACTT GCTATCTGAT TTCGATTAAA
151 ACTGGTGGAT GCGTTGAAGA TTGCGCCTAC TGTGCCCAAT CTCCCCGCTA
201 TCATACCCAC GTCACACCAG AACCTATGAT GAAAATTGTA GACGTTGTGG
251 AAAGGGCAAA ACGTGCTGTA GAGCTAGGCG CCACTCGTGT GTGCTTTGGG
301 GCTGCCTGGC GCAATGCTAA GGACGATCGA TACTTTGATA GAGTCCTCGC
351 TATGGTGAAG AGTATCACAG ATCTCGGAGC CGAGGTTTGT TGTGCTTTAG
401 GCATGCTCTC CGAAGAGCAA GCTAAAAAAC TGTATGATGC AGGACTTTAT
451 GCCTACAATC ATAATTTAGA CTCTTCTCCG GAATTCATATG AAACATAAAT
501 CACAACACGT TCTTATGAAG ATCGCCTCAA CACTCTTGAT GTAGTAAATA
551 AATCTGGCAT TAGTACATGC TCGGGTGGTA TTGTAGGTAT GGGAGAATCT
601 GAAGAAGACC GTATAAAGCT TCTTCATGTT CTTGCAACAA GAGATCATAT
651 CCCAGAATCC GTACCTGTAA ATTTACTTTG GCCGATTGAC GGCACGCCTT
701 TGCAAGACCA GCCTCCGATT TCTTCTGGG AAGTCTTGCG AACCATAGCA
751 ACGGCACGGG TTGTTTTCCC CAGATCCATG GTACGACTTG CTGCAGGACG
801 CGCTTTCCTC ACAGTAGAAC AACAAACCTT ATGTTTTCTA GCCGGTGCCA
851 ACTCCATATT CTATGGAGAT AAACGTGTTA CTGTAGAAAA CAATGATATA
901 GATGAAGATG CTGAAATGAT CAAACTTTTA GGCTTAATCC CTCGCCCTTC
951 ATTTGGAATA GAAAGAGGTA ACCCATGTTA TGCCAACAAT TCCTAA

```

The PSORT algorithm predicts cytoplasm (0.2926).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 142A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 142B) and for FACS analysis.

35 These experiments show that cp7377 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

**Example 143**

The following *C.pneumoniae* protein (PID 4377407) was expressed <SEQ ID 285; cp7407>:

```

40 1  MVECPNNSWFR MCGNFNCEWV EVTTTEETTR QSASDISEEA GSSGGAAPIT
51  TQPTKITKVE KRVQFNTAQG DESTIHMIE AGELVDSILS HRRTQGCTEY
101 CYDSYATGCG QRCGSFGRLI CGTYKACCLD REDNQVAGLV HECEQTHGPI
151 AVALAAKTMG LNLMELEVEKN TILSEEQKNE FRQHCSEAKT QLYGTMQSLS
201 QNFFLEGVNS IRERGLDDSL VQAVLSFIAT RSWEKTIESE EASGTSSASN
45 251 STRIPACYIL NTSPLTTSRL SCGSRDARRP SSVGAEPQYV AKKYNDNGMA
301 RQLGKIQVTN LKTGDFSALG PFGLLIVKML NSFLLSASQS TSSILKHTGG
351 EICYTCPNFR DIVVLLMLAI GYCPANTDET SVVDIHMIDD PIMTIFYRLQ
401 YSYRTGKTS SFLKKKPSLV RQESLDCPTP AESVPLMSSL EEEDENEDDD
451 EDGNLAYQQR ILECSGHLQT LFLGIKINKE *

```

The cp7407 nucleotide sequence <SEQ ID 286> is:

```

50 1  ATGGTTTGCC CAAATAATTC TTGGTTCAGA ATGTGTGGAA ATTTCAACTG
51  CGAATGGGTT GAAGTAACAA CAACAGAAGA AACACGCGG CAATCGGCTT
101 CAGATATAAG CGAAGAAGCT GGTTCGAGTG GAGGAGCTGC TCCTATAACT
151 ACGCAACCTA CTAAAATTAC AAAAGTAGAG AAACGTGTCC AATTTAATAC

```

The PSORT algorithm predicts inner membrane (0.7453).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 140A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 140B) and for FACS analysis.

- 5 These experiments show that cp7359 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 141

The following *C.pneumoniae* protein (PID 4377374) was expressed <SEQ ID 281; cp7374>:

```

10      1  MDKQSSGNSG CIWHPFTQSA LDSTPIKIVR GEGAYLYAES GTRYLDAISS
      51  WWCNLHGHGH PYITKKLCEQ AQKLEHVIFA NFTPHEPALEL VSKLAPLLPE
     101  GLERFFFSND GSTSIEIAMK IAVQYYYNQN KAKSHFVGLS NAYHGDTFGA
     151  MSIAGTSPTT VPFHDLFLPS STIAAPYYGK EELAIQAQAKT VFSIESNIAAF
     201  IYEPLLQAGG GMLMYNPEGL KEILKLAKHY GVLICIADEIL TGFGRTGPLF
     251  ASEFTDIPPD IICLSKGLTG GYLPLALTVT TKEIHDAFVS QDRMKALLHG
15      301  HTFTGNPLGC SAALASLDLT LSPECLQQRQ MIERCHQEFQ EAHGSLWQRC
     351  EVLGTVLALD YPAEATGYFS QYRDHLNRFF LERGVLLRPL GNTLYVLPYP
     401  CIQEEDLRRI YSHLQDALCL QPQ*
```

The cp7374 nucleotide sequence <SEQ ID 282> is:

```

20      1  ATGGACAAGC AATCATCAGG GAATTCAGGG TGTATCTGGC ACCCCTTCAC
      51  TCAATCTGCA TTAGATTCTA CACCCATAAA GATTGTAAGG GGAGAAGGTG
     101  CTTACCTCTA TCGCGAATCA GGAACAAGAT ATCTTGATGC GATATCTTCA
     151  TGGTGGTGCA ACCTCCACGG TCATGGGCAT CCCTACATTA CAAAAAATT
     201  ATGTGAGCAA GCACAGAAGT TAGAACATGT GATCTTCGCA AATTTCACCC
     251  ATGAACCGGC TCTAGAGCTC GTATCGAAAC TCGCTCCCCT CCTTCCTGAA
25      301  GGTCTAGAAC GTTCTTTTTT CTCTGACAAC GGATCAACGT CTATCGAAAT
     351  AGCAATGAAA ATTGCTGTGC AATATTACTA CAATCAAAAC AAGGCTAAGA
     401  GCCATTTTGT TGGACTCAGC AATGCCTATC ACGGAGATAC ATTTGGAGCT
     451  ATGTCGATAG CTGGCACGAG CCCTACTACA GTTCCCTTTC ATGATCTTTT
     501  TCTTCTTCC AGTACAATTG CTGCTCCCTA TTATGGCAAG GAAGAGCTTG
30      551  CCATTGCCCA AGCAAAAACA GTCTTTTCTG AAAGCAATAT CGCAGCGTTT
     601  ATCTATGAGC CGCTATTGCA AGGTGCTGGA GGGATGTTAA TGTATAATCC
     651  CGAAGGCCTA AAGGAGATTC TCAAGCTTGC CAAGCATTAC GGGGTCTCTT
     701  GTATGTGCTG TGAATTTCTT ACTGGCTTTG GCCGTACGGG TCCACTGTTT
     751  GCTTCTGAAT TTACAGACAT TCCTCCTGAC ATTATCTGTC TTCTTAAAGG
35      801  TCTTACAGGA GGCTATCTCC CTCTAGCCTT GACAGTAACC ACTAAAGAAA
     851  TTCTGATGTC CTTGTCTCC CAAGATCGGA TGAAGGCACT GCTTCTGTC
     901  CATACCTTCA CAGGAAATCC TTAGGCTGT AGTGCTGCCC TCGCTTCTTT
     951  GGATCTCACC CTATCTCCAG AATGCCTACA ACAAAGGCAA ATGATAGAAC
40     1001  GGTGTCATCA AGAGTTTCAA GAAGCTCATG GTTCCCTATG GCAACGGTGT
     1051  GAGGTTCTGG GCACGGTACT CGCTCTAGAT TACCCTGCAG AAGCTACAGG
     1101  ATATTTTTC CAATATAGAG ACCATCTCAA TCGCTTTTTC TTAGAACGTG
     1151  GAGTCCTTCT TCGTCCTTTA GGAACACAC TGTATGTGCT GCCCCCTTAC
     1201  TGTATCCAAG AAGAAGATCT CCGGATTATT TATCTCACC TACAGGATGC
     1251  CCTATGTCTA CAACCACAGT AA
```

- 45 The PSORT algorithm predicts cytoplasm (0.2930).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 141A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 141B) and for FACS analysis.

- 50 These experiments show that cp7374 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

These experiments show that cp6432 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 145

The following *C.pneumoniae* protein (PID 4376433) was expressed <SEQ ID 289; cp6433>:

```

5      1  MNWVPKTI DH VDPESIDIR KVVSCYKLIK ECQPEFRSLI SELLGVIRCG
      51  LRLLKRSKYQ EQARTVSDDE APLFCLTRSY YQDGYLTPLR AGPRDLINHY
     101  IHLRRRENPK HFFSPKHPCY YARLAFNESV CVYRELFIDIE RLTKMYVEGD
     151  YSKEQEKNLQ AILSFVKTL D EGKDFLIEHK DTDLIGRGFT DVFCT*
  
```

The cp6433 nucleotide sequence <SEQ ID 290> is:

```

10      1  ATGAATTGGG TTCCAAAAAC AATAGACCAT GTAGATCCAG AATCAGAGAT
      51  AGATATACGT AAAGTCGTCT CCTGCTATAA GTTGATAAAA GAATGTCAAC
     101  CTGAATTTTCG ATCTCTTATA AGTGAATTAC TAGGAGTGAT TCGGTGTGGC
     151  TTAAGACTAT TAAAACGTTT TAAGTATCAA GAACAGGCTA GAACTGTATC
     201  TGATGAAGAT GCACCTCTTT TCTGCCTGAC TCGTTCCTAT TATCAAGATG
     15  251  GTTATCTCAC GCCATTAAGA GCAGGACCTC GTGATCTTAT AAATCACTAT
      301  ATACACTTGC GTCGCCGAGA GAATCCTAAG CATTTTTCAT GTCCTAAGCA
      351  TCCATGTTAT TATGCTCGAT TGGCTTTTAA TGAGTCAGTG TGTGTCTATA
     401  GAGAACTCTT TGATATAGAG CGACTTACAA AAATGTATGT CGAGGGTGAT
     451  TATTCTAAAG AACAAGAGAA AAACCTACAG GCTATTCTTA GTTTTGTGAA
     20  501  AACTCTAGAT GAAGGAAAGG ACTTTCTTAT TGAACATAAA GATACCGATC
     551  TCATTGGGAG AGGTTTTACT GATGTGTTCT GCACTTAA
  
```

The PSORT algorithm predicts cytoplasm (0.4068).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 145A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 145B) and for FACS analysis.

These experiments show that cp6433 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 146

The following *C.pneumoniae* protein (PID 4376643) was expressed <SEQ ID 291; cp6643>:

```

30      1  MGYLPVSATD VLFESPAAPL INSANTQNQK LIELKGKQQA ESSPRITTSV
      51  ILEVLLVIGC CLIVLSLLAI RPALQFTLET GHPAIAVLA VSGTILLVAV
     101  IILFCFLAAV PFAAKKTYKY VKTVDDYASW HSHQQTPTLG TIFSGIVYAE
     151  SQAQL*
  
```

The cp6643 nucleotide sequence <SEQ ID 292> is:

```

35      1  ATGGGATATC TTCCAGTATC TGCTACGGAC GTTCTTTTTCG AAAGTCCAGC
      51  CGCTCCCTTA ATCAATAGCG CAAACACACA AAATCAGAAA CTCATAGAAC
     101  TCAAGGGGAA GCAGCAAGCT GAGTCTTCTC CACGGACAAT CACTTCTGTC
     151  ATATTGGAAG TTCTCCTAGT GATCGGATGC TGCCTCATAG TTCTTAGTTT
     201  ATTGGCAATC CGCCCTGCTC TGCAATTAC TCTAGAACT GGACATCCAG
     40  251  CTGCCATTGC AGTCCTTGCT GTCTCAGGAA CAATTCTATT GGTGGCTGTT
      301  ATCATCTTGT TTTGCTTTCT AGCAGCTGTG CCATTCGCTG CTAAGAAAAC
      351  TTATAAATAT GTTAAGACGG TTGATGACTA TGCTTCTTGG CATTCTCATC
     401  AGCAAACACC GACCCTAGGC ACTATCTTTT CAGGTATCGT CTATGCAGAA
     451  TCCCAGGCGC AATTATAG
  
```

45 The PSORT algorithm predicts inner membrane (0.6859).

201 TGCTCAAGGT GATGAAAGTA CAATACACAT GATCCAAGAA GCAGGAGAAT  
 251 TGGTAGACTC CATTCTATCA CATAGACGAA CGCAAGGATG TACAGAGTAT  
 301 TGTTATGACA GTTACGCAAC TGGATGTGGT CAGCGTTGCG GATCTTTTGG  
 351 AAGACTCATT TGTGGAACGT ATAAAGCGTG TTGCTTAGAC AGAGAGGATA  
 5 401 ATCAGGTTGC TGGACTTGTC CATGAATGCG AACAGACCCA TGGTCCTATT  
 451 GCCGTTGCTT TAGCTGCTAA AACTATGGGC CTCAACTTAA TGGAACTTGT  
 501 AGAAAAAAC ACTATTTTGT CTGAAGAAC GAAAAATGAA TTTAGACAGC  
 551 ATTGCTCGGA AGCTAAACC CAACTCTATG GAACGATGCA GAGCCTTTCT  
 601 CAAAACTTT TCCTTGAAGG AGTCAACAGC ATTAGAGAAC GCGGTCTAGA  
 10 651 CGATTCACTA GTCCAAGCCG TGCTAAGCTT TATTGCTACA AGGTCTTGGG  
 701 AAAAACTAT AGAATCAGAG GAAGCCTCAG GAACATCTTC TGCTTCTAAT  
 751 TCTACACGCA TTCTTGCGTG CTATATCTTA AATACGAGCC CCTTAACGAC  
 801 GTCACGCGTA TCCTGTGGAT CAAGAGATGC GCGACGCCCA TCTTCAGTCG  
 851 GTGCAGAGCC CCAGTACGTA GCAAAAAAAT ACAATGACAA TGGCATGGCC  
 15 901 AGACAATTAG GAAAAATCCA AGTCACCAAT CTAAAAACAG GAGATTTTTC  
 951 AGCTTTTAGGT CCTTTTGGTC TCCTGATTGT GAAAATGCTG AATAGCTTTC  
 1001 TCTTATCTGC ATCACAAGC ACATCTTCTA TTCTAAAGCA CACAGGTGGA  
 1051 GAAATATGTT ATACGTGCCC AAATTTTCGT GATATCGTCG TTTTATTGAT  
 1101 GTTAGCGATT GGCTATTGCC CTGCAAATAC CGATGAGACA TCTGTCGTAG  
 20 1151 ATATACACAT GATAGATGAT CCGATTATGA CCATCTTCTA TCGACTACAA  
 1201 TACAGCTATA GAACAGGGAA AACTTCAGCA TCGTTTTTAA AAAAGAAACC  
 1251 CTCATTAGTA AGACAGGAAA GTCTTGATTG TCCTACCCCT GCAGAATCTG  
 1301 TCCCTCTCAT GTCAAGTCTC GAAGAAGAAG ATGAAAATGA AGATGATGAT  
 1351 GAGGATGGGA ATTTGGCGTA TCAACAGCGT ATCCTTGAAT GCTCGGGTCA  
 25 1401 TTTACAACT CTATTTTATG GGATAAAAT AAACAAAGAA TAA

The PSORT algorithm predicts inner membrane (0.1319).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 143A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 143B) and for FACS analysis.

30 These experiments show that cp7407 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone:

#### Example 144

The following *C.pneumoniae* protein (PID 4376432) was expressed <SEQ ID 287; cp6432>:

35 1 MTRSTIESSD SLCSRSFSQK LSVQTLKNLC ESRLMKITSL VIAFLTLLVG  
 51 GALIALAGGG VLSFPLGLIL GSVLVLFSSI YLVSCCKFFT LKEMTMTCVS  
 101 KSKINIWF EK QRMKIDIEAL ENPDFGENK RNVGNRSARN QLEMILHETD  
 151 GIILKRYMKG AKMYFYL\*

The cp6432 nucleotide sequence <SEQ ID 288> is:

40 1 ATGACTAGAA GTACTATTGA AAGCAGTGAT TCGCTATGCT CAAGGTCTTT  
 51 TTCTCAAAAA TTAAGTGTCC AGACATTAAA AAATCTCTGT GAAAGTAGAT  
 101 TAATGAAGAT CACTTCTCTT GTGATTGCTT TCCTAACTCT AATTGTGGGG  
 151 GGTGCTCTTA TAGCTTTAGC AGGAGGGGGG GTTCTTTCTT TCCCTCTTGG  
 201 GCTAATCTTA GGAAGCGTAC TCGTTTGTGTT TTCTTCTATC TATTTAGTCT  
 251 CTTGTTGTAA ATTTTCTACT TTAAGAGAGA TGACAATGAC CTGTAGTGTC  
 45 301 AAATCTAAAA TCAATATATG GTTTGAAAAAG CAACGAAACA AAGACATCGA  
 351 AAAGGCATTA GAGAATCCAG ATCTCTTTGG AGAAAATAAG AGAAATGTTG  
 401 GAAATCGTTC GGCAAGAAAT CAACTAGAAA TGATCTTACA CGAGACTGAC  
 451 GGAATTATTT TGAAAAGATA TATGAAAGGA GCTAAAATGT ACTTTTATTT  
 501 ATGA

50 The PSORT algorithm predicts inner membrane (0.5394).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 144A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 144B) and for FACS analysis.

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 148A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 148B) and for FACS analysis.

These experiments show that cp7253 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 149

The following *C.pneumoniae* protein (PID 4376264) was expressed <SEQ ID 297; cp6264>:

```

1  VISGLLFLLV RREVPTVRSE EIPRGVSVTP SEEPALEKAQ KEPETKKILD
51  RLPKELDQLD TYIQEVFACL ERLKDPKYED RGLLTEAKEK LRVFDVVEKD
101 MMSEFLDIQR VLNEEAYYVE HCQDPLENIA YEIFSSQELR DYICAGVCGY
151 LPSGDARADR LKRSVKEVMD RFMRVTWKSX EASVMLDHSY GVARELFKKA
201 VGVLEESVYK ILFKSYRDAF YECEKAKIQR DGRFKWL*
```

The cp6264 nucleotide sequence <SEQ ID 298> is:

```

1  GTGATTTCGG GACTTCTATT CCTTCTAGTA AGACGAGAGG TTCCGACAGT
51  ACGTTCAGAG GAAATTCCCA GAGGGGTTTC TGTGACCCCT TCTGAAGAGC
101 CTGCTCTAGA GAAGGCTCAA AAAGAACCGG AGACAAAGAA AATTTTAGAT
151 CGGTTGCCGA AGGAATTGGA TCAGTTAGAT ACGTATATTC AGGAAGTGTT
201 TGCATGTTTA GAGAGGCTGA AGGATCCTAA GTACGAAGAT CGAGGTCCTT
251 TAACAGAGGC GAAGGAGAAA CTTCGAGTTT TTGACGTTGT TGAGAAAGAT
301 ATGATGTCAG AGTTTTTAGA CATAACGA GTGTTGAATG AGGAAGCATA
351 TTATGTAGAA CATTGTCAAG ATCCCCTAGA GAATATAGCC TACGAGATTT
401 TCTCTTCCCA AGAGCTTCGT GATTACTACT GTGCAGGGGT GTGTGGGTAT
451 TTGCCTTCG GGGATGCTCG AGCGGATCGA TTAAAGAGAT CAGTTAAGGA
501 GGTAATGGAT CGCTTTATGA GGGTGACCTG GAAATCTTGG GAGGCATCAG
551 TCATGTTGGA TCATAGCTAT GGGGTAGCGC GAGAGTTATT CAAGAAGGCA
601 GTAGGAGTAC TAGAGGAGAG TGTCTATAAA ATTCTGTTTA AGAGCTATAG
651 AGATGCGTTT TATGAATGTG AGAAGGCAA GATCCAGAGG GATGGGCGTT
701 TCAAATGGTT ATAG
```

The PSORT algorithm predicts cytoplasm (0.2817).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 149A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 149B) and for FACS analysis.

These experiments show that cp6264 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 150

The following *C.pneumoniae* protein (PID 4376266) was expressed <SEQ ID 299; cp6266>:

```

1  MLLISGALF LTLGIPGLSA AISFGLGIGL SALGGVLMIS GLLCLLVKRE
51  IPTVRPEEIP EGVSLAPSEE PALQAAQKTL AQLPKELDQL DTDIQEVFAC
101 LRKLKDSKYE SRSFLNDAK ELRVFDFVVE DTLSEIFELR QIVAQEGWDL
151 NFLINGGRSL MMTAESESLD LFHVSKRLGY LPSGDVRGEG LKKSACEIVA
201 RLMSLHCEIH KVAVAFDRNS YAMAEKAFK ALGALEESVY RSLTQSYRDK
251 FLESERAKIP WNGHITWLRL DAKSGCAEKK LGMPRNVGRN LGKQSFQ*
```

The cp6266 nucleotide sequence <SEQ ID 300> is:

```

1  ATGCTCTTAC TGATTTCAGG AGCTCTCTTT CTGACGTTAG GGATTCCAGG
51  ATTGATGCA GCAATTTCTT TTGGATTAGG CATCGGTCTC TCCGCATTAG
101 GAGGAGTGCT GATGATTTTC GGAATACTAT GTCTTTTAGT AAAACGAGAG
151 ATTCCGACAG TACGACCAGA AGAAATTCCT GAAGGGGTTT CGCTGGCTCC
```

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 146A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 146B) and for FACS analysis.

These experiments show that cp6643 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 147

The following *C.pneumoniae* protein (PID 4376722) was expressed <SEQ ID 293; cp6722>:

```

1  VSSTLNGVFP SSLPEESADL FITNKEIVAL GEKGNVFLTH SIPMHIAAIT
51  ILVIVALAGI AIICLGCSYQ SILLIAVGIV LTILTLLCLQ ALVGFIKFIK
101 QLPQQLHTTV QFIREKIRPE SSLQLVTNAQ RKTQDTLKL YEELCDLSQK
151 EFKLQSTLYQ KRFELSHKNE KTNQN*

```

The cp6722 nucleotide sequence <SEQ ID 294> is:

```

1  GTGTCTAGTA CTTTAAACGG GGTATTTCCC TCATCCCTTC CGGAAGAGTC
51  TGCTGATTTA TTCATTACGA ATAAGGAGAT CGTAGCTTTG GGGGAGAAGG
101 GCAATGTTTT TCTCACCAC TCCATTCTTA TGCATATTGC TGCGATTACG
151 ATCTTAGTGA TTGTAGCTCT TGCTGGAATC GCTATTATCT GTTTGGGTTG
201 CTATAGCCAA AGCATTCTGT TGATTGCCGT TGGCATTGTT CTTACTATTT
251 TGACTCTTCT CTGCCTACAA GCCTTGCTAG GATTTATTAA ATTCATCCGG
301 CAGCTCCCTC AGCAGCTCCA TACGACAGTA CAATTTATCA GGGAGAAGAT
351 TCGACCTGAA TCCTCTCTAC AGCTTGTAAC CAATGCACAG AGAAAAACCA
401 CTCAAGATAC GCTAAAGTTA TACGAAGAAC TCTGCGACCT CTCACAAAAA
451 GAGTTCAAAC TGCAATCAAC TCTTTATCAA AAACGTTTTG AGCTTTCTCA
501 CAAGAATGAA AAGACAAATC AAAACTAG

```

The PSORT algorithm predicts inner membrane (0.6668).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 147A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 147B) and for FACS analysis.

These experiments show that cp6722 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 148

The following *C.pneumoniae* protein (PID 4377253) was expressed <SEQ ID 295; cp7253>:

```

1  MSELAPCSTG LQMVPHTQVH HALDTRRVIL TIAACLSLIA GIVLVGLGAA
51  AILPSLFGVI GGMILILFSS IALIYLYKKT REVDQIALEP LPEMISKDQS
101 IIDFVKTRDY ASLEKKATFA YTHTHYDGS MVFYREIPRF MLGSYLALRK
151 DMDRQALF*

```

The cp7253 nucleotide sequence <SEQ ID 296> is:

```

1  ATGAGCGAGC TCGCCCCCTG CTCGACAGGA TTGCAGATGG TCCCCATAC
51  GCAGGTCCAT CATGCCCTTG ATACGCGGAG AGTCATTCTA ACGATAGCCG
101 CCTGTCTGTC TTTAATTGCA GGAATCGTGT TGGTTGGCTT AGGTGCTGCA
151 GCAATCCTGC CCTCGCTTTT TGGAGTCATT GGAGGAATGA TTCTTATTCT
201 GTTTCTTTCG ATCGCCCTCA TTTATTATTA CAAGAAGACA AGGGAGGTGG
251 ATCAGATTGC TCTGGAGCCT CTTCTGAGA TGATTCTTAA AGATCAAAGC
301 ATTATAGATT TTGTAAAGAC ACGAGACTAT GCATCTTTAG AAAAGAAAGC
351 GACCTTTGCT TATACTCATA CTCATTATTA CGATGGAAGC ATGGTCTTCT
401 ATAGGGAGAT CCCTAGATTT ATGTTAGGCT CTTATCTCGC GCTTCGCAA
451 GACATGGACC GCCAAGCTCT TTTTGA

```

The PSORT algorithm predicts inner membrane (0.5394).

The cp6282 nucleotide sequence <SEQ ID 304> is:

```

1  ATGTCCTTAT TGAACCTTCC CTCAAGCCAG GATTCTGCAT CTGAGGACTC
51 CACATCGCAA TCTCAAATCT TCGATCCCAT TAGAAATCGG GAGTTAGTTT
101 CTACTCCCGA AGAAAAAGTC CGCCAAAGGT TGCTCTCCTT CCTAATGCAT
5  151 AAGCTGAACT ACCCTAAGAA ACTCATCATC ATAGAAAAAG AACTCAAAAC
201 TCTTTTTCCT CTGCTTATGC GTAAAGGAAC CCTAATCCCA AAACGCCGCC
251 CAGATATTCT CATCATCACT CCCCCACAT ACACAGACGC ACAGGGAAAC
301 ACTCACAACC TAGGCGACCC AAAACCCCTG CTACTTATCG AATGTAAGGC
351 CTTAGCCGTA AACCAAAATG CACTCAAACA ACTCCTTAGC TATAACTACT
10 401 CTATCGGAGC CACCTGCATT GCTATGGCAG GGAAACACTC TCAAGTGTC
451 GCTCTCTTCA ATCCAAAAC ACAAACCTCT GATTTTATC CTGGCCTCCC
501 AGAGTATTC CAACTCCTAA ACTACTTTAT TTCTTTAAAC TTATAG

```

The PSORT algorithm predicts cytoplasm (0.362).

The following *C.pneumoniae* protein (PID 4377373) was also expressed <SEQ ID 305; cp7373>:

```

15 1  MSTTTVKHFI HTASRWEFVL KEIVASNYWH AQWINTLSFL ENSGAKKISA
51  SEHPTEVKEE VLKHAEEFR HGHYLKTQIS RISETSLPDY TSKNLLGGLL
101 TKYYLHLLDL RTCRVLENEY SLGQTLKTA AYILVTYAIE LRASELYPLY
151 HDILKEAQSK ITVKSIILEE QGHLQEMERE LKDLPHGEEL LGYACQFEGE
201 LCLQFVERLE QMIFDPSTF TKF*

```

20 The cp7373 nucleotide sequence <SEQ ID 306> is:

```

1  ATGTCTACAA CCACAGTAAA AACTTTTATC CACACAGCCT CTCGTTGGGA
51  GCCCGTTCTC AAAGAGATCG TAGCTTCCAA CTATTGGCAT GCACAATGGA
101 TAAATACCCT GTCCTTTTTA GAAAATAGTG GAGCAAAAAA AATCTCCGCA
25  151 AGTGAACATC CTACGGAGGT AAAGGAAGAA GTTTTAAAC ATGCTGCTGA
201 AGAATTTTCGT CATGGTCACT ATCTAAAAAC TCAGATTTCT AGAATCTCAG
251 AGACTTCTCT CCCTGACTAT ACATCTAAAA ATCTTCTGGG AGGCTTACTT
301 ACAAATATAT ACCTCCATCT TCTAGATTTA AGGACGTGCC GAGTACTGGA
351 AAATGAATAC TCCCTATCGG GACAAACGTT AAAAAGTGCA GCGTATATTT
401 TAGTTACCTA CGCAATCGAA CTTTCGTGCTT CTGAACTTTA TCCTCTGTAT
30 451 CACGATATTC TGAAAGAAGC TCAAAGTAAA ATAACGGTAA AATCCATTAT
501 CTTAGAAGAG CAAGGCCATC TGCAAGAGAT GGAACGTGAA CTTAAAGATC
551 TCCCCACGG GGAGGAAGTC TTAGGCTATG CTTGCCAATT CGAAGGGGAG
601 CTTTGCTTGC AGTTTGTAGA GAGATTAGAA CAAATGATCT TCGATCCTTC
651 CTCGACTTTT ACAAAGTTCT AG

```

35 The PSORT algorithm predicts cytoplasm (0.1069).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 152A; 6282 = lanes 8 & 9; 7373 = lanes 2-4). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 152B & 153) and for FACS analysis.

40 These experiments show that cp6282 & cp7373 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequence alone.

Example 154 ,  
 Example 155 ,  
 Example 156 ,  
 Example 157 and  
 45 Example 158

The following *C.pneumoniae* protein (PID 4376412) was expressed <SEQ ID 307; cp6412>:

```

50 1  MSSSEVVVFQT VHGLGFGLS SKSVVPFKKS LSDAPRVVCS ILVLTGLGLA
51  LVCGIAITCW CVPGVILMGG ICAIVLGAIS LALSFLWLWG LPSNCCGSKR
101 VLPGEGLLRD KLLDGGFSRA APSGMGLPGD GSPRASTPSC LEELQAEIQA
151 VTQAIDQMSD D*

```

The cp6412 nucleotide sequence <SEQ ID 308> is:



-169-

```

201 TTCTGAGGAG CCAGCTCTAC AGGCAGCTCA GAAGACTTTA GCTCAGCTGC
251 CTAAGGAATT GGATCAGTTA GATACAGATA TTCAGGAAGT GTTCGCATGT
301 TTAAGAAAGC TGAAAGATTC TAAGTATGAA AGTCGAAGTT TTTTAAACGA
351 TGCTAAGAAG GAGCTTCGAG TTTTGTGACTT TGTGGTTGAG GATACCTCT
401 CGGAGATTTT CGAGTTGCGG CAGATTGTGG CTCAAGAGGG ATGGGATTTA
451 AACTTTTGA TCAATGGGGG ACGAAGCCTC ATGATGACTG CAGAATCTGA
501 ATCGCTTGAT TTGTTTCATG TATCGAAGCG GCTAGGGTAT TTACCTTCTG
551 GGGATGTTTC AGGGGAGGGG TTAAAGAAAT CTGCGAAGGA GATAGTCGCT
601 CGTTTGATGA GCTTGCATTG CGAGATTCAC AAGGTGGCGG TAGCGTTTGA
651 TAGGAATTCC TATGCGATGG CAGAAAAGGC GTTTGCGAAA GCGTTGGGAG
701 CTTTAGAAGA GAGTGTGTAT CGGAGTCTGA CGCAGAGTTA TAGAGATAAA
751 TTTTGGGAGA GCGAGAGGGC GAAGATCCCA TGGAATGGGC ATATAACCTG
801 GTTAAGAGAT GATGCGAAGA GTGGGTGTGC TGAAAAGAAG CTCGGGATGC
851 CGAGGAACGT TGGAAGAAAT TTAGGAAAGC AGTCTTTTGG GTAG

```

15 The PSORT algorithm predicts inner membrane (0.3590).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 150A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 150) and for FACS analysis.

20 These experiments show that cp6266 is a surface-exposed and immunoaccessible protein and that they it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 151

The following *C.pneumoniae* protein (PID 4376895) was expressed <SEQ ID 301; cp6895>:

```

1 MKIKKSFQYS LCQAKRFQNM LPNHFDPCLO PVNLQLKQDR LAYGELIILL
25 51 SKYQQKTFSS LLKEETCSLN RAKQHLLYKI LRDFNTMQHL RSLGLNGWGE
101 IPMSPC*

```

The cp6895 nucleotide sequence <SEQ ID 302> is:

```

1 ATGAAGATTA AAAAATCTTT TCAATACAGT TTATGCCAAG CAAAGAGATT
25 51 TCAGAACATG CTGCCAAACC ACTTTGATCC ATGTTTGCAG CCAGTGAATT
30 101 TACAACTCAA ACAAGACAGA TTGGCATACG GGGAGCTCAT CATATTGCTA
151 TCTAAATATC AACAAAAGAC CTTTTCCTCT TTGTTGAAGG AAGAAACATG
201 TTCTCTTAAT CGTGCGAAGC AGCACTTATT GTATAAGATT TTGAGAGATT
251 TTAATACTAT GCAGCATCTA AGGTCCCTCG GATTAAATGG TTGGGGAGAG
301 ATCCCTATGA GTCCTTGCCT CTAA

```

The PSORT algorithm predicts cytoplasm (0.3264).

35 The protein was expressed in *E.coli* and purified as a his-tag product (Figure 151A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 151B) and for FACS analysis.

These experiments show that cp6895 is a surface-exposed and immunoaccessible protein and that it is a useful immunogen. These properties are not evident from the sequence alone.

### 40 Example 152 and Example 153

The following *C.pneumoniae* protein (PID 4376282) was expressed <SEQ ID 303; cp6282>:

```

1 MSLLNLPSSQ DSASEDSTSQ SQIFDPIRNR ELVSTPEEKV RQRLLSFLMH
45 51 KLNYPKLLII IEKELKTLFP LLMRKGTLP KRRPDILIT PPTYTDAQGN
101 THNLGDPKPL LLIECKALAV NQNALKQLLS YNYSIGATCI AMAGKHSQVS
151 ALFNPKTQTL DFYPGLPEYS QLLNYFISLN L*

```

The PSORT algorithm predicts inner membrane (0.5989).

The following *C.pneumoniae* protein (PID 4376654) was also expressed <SEQ ID 315; cp6654>:

```

1  MKTKMNSRKK AGQWAFNSP TPGVSSTLVL AWTWPWGYDK DVQDILERKD
51 PMSSSLSEKD SKEFLKNLFV DLLENGFTSV HIHAEFAFTP LDHTGKPHFK
101 RDNVYLPGLK LGALNEAAVQ ANVSADTQFT LFLTQDECNP FHDKKRG*

```

The cp6654 nucleotide sequence <SEQ ID 316> is:

```

1  ATGAAACTA AAATGAACTC TAGAAAAAAA GCAGGTCAAT GGGCAATTTT
51 CAATTCTCCA ACTCCTGGTG TCAGTTCAAC TTTAGTTTTA GCATGGACTC
101 CTTGGGGTTA TTACGACAAG GATGTACAAG ATATCTTAGA AAGAAAAGAT
151 CCGATGAGCT CTTTCGCTTTC TGAAAAAGAC TCAAAGGAGT TCTTGAAAAA
201 TCTGTTTGTA GATCTCTTAG AAAATGGCTT CACATCAGTA CATATTCACG
251 CAGAAGAAGC TTTCACCTCT CTTGATCATA CCGGGAAACC TCACTTTAAA
301 AGAGACAATG TGTAATACCG CGGAAAGTTG TTAGGCGCCT TGAATGAGGC
351 TGCGGTACAA GCCAATGTAA GTGCGGATAC TCAATTACAA TTGTTCTCTT
401 CTAAGATGA GTGCAATCCT TTTTCATGATA AGAAAAGAGG TTAA

```

The PSORT algorithm predicts cytoplasm (0.0730).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 154A; 6412 = lanes 2-3; 6431 = lanes 11-12; 6443 = lanes 5-6; 6496 = lanes 8-9; 6654 = lane 10; markers in lanes 1, 4, 7). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 154B, 155, 156, 157 & 158) and for FACS analysis.

These experiments show that cp6412, cp6431, cp6443, cp6496 & cp6654 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from their sequences alone.

#### Example 159 and Example 160

The following *C.pneumoniae* protein (PID 4376477) was expressed <SEQ ID 317; cp6477>:

```

1  LLKFFLVCEE LCILTVATHR ALLETPLALS FFKELKTKYV YRAKDILQLH
51 NYKGFTILNT SPLCS*

```

The cp6477 nucleotide sequence <SEQ ID 318> is:

```

1  TTGCTAAAGT TCTTTCTAGT ATGTGAAGAG TTATGTATAC TTACTGTTGC
51 TACACATAGA GCTCTCTTAG AAACCTCCTT AGCTCTATCA TTTTAAAG
101 AACTTAAGAC AAAATATGTC TACAGGGCGA AAGACATACT ACAACTACAT
151 AACTATAAAG GATTACTAT CCTTAATACA TCACCGTTAT GTTCTTAA

```

The PSORT algorithm predicts inner membrane (0.128).

The following *C.pneumoniae* protein (PID 4376435) was also expressed <SEQ ID 319; cp6435>:

```

1  LWSHFPRGFF MLPFCPTILL AKPFLNSEN YGLERLAATVD SYFDLGQSQI
51 VFLSKQDQGI TVEELSAKDR KFKPGSMNCT LYTEDPILPA HNSFSNCSDI
101 QMRTPISPIH *

```

The cp6435 nucleotide sequence <SEQ ID 320> is:

```

1  TTGTGGTCGC ATTTCCCAAG AGGATTTTTT ATGCTCCCTT TTTGCCCTAC
51 CATCCTTCTT GCTAAACCTT TTTTAAATAG CGAGAATTAC GGCTTAGAAC
101 GTTTAGCTGC AACCGTAGAT TCTTATTTTG ATCTGGGACA GTCTCAAATA
151 GTCTTCCTAA GCAAACAGGA TCAAGGAATC ACTGTGGAAG AATTGAGTGC
201 TAAAGATAGG AAATTCAAGC CAGGCTCTAT GAACTGTACA CTGTACACTG
251 AAGATCCTAT CTTACCTGCT CATAATTCCT TTAGTAATTG CTCTGATATT
301 CAAATGCGTA CTCCGATTAG CCCTATACAT TAA

```

-171-

1 ATGAGCAGTT CGGAAGTTGT TTTCCAGACA GTTCATGGCC TTGGCTTTGG  
 51 TGGATTGTCT TCAAAAAGTG TTGTCCCTTT TAAGAAAAGT CTTTCGGATG  
 101 CGCCCCGTGT TGTGTGCTCG ATTTTAGTTT TGACTCTGGG GTTGGGAGCG  
 151 CTTGTTTGTG GTATTGCCAT TACTTGTGG TGTGTCCCGG GAGTTATTTT  
 5 201 AATGGGGGGA ATTTGCGCTA TAGTTTATAG TGCAATTCT TTAGCTTTAA  
 251 GTCATATTTG GTTGTGGGGT TTATTTTCTA ATTGTTGTGG TTCTAAGAGA  
 301 GTTTTACCGG GTGAGGGATT GCTACGGGAT AAGCTTTTAG ATGGTGGATT  
 351 TTCAAGAGCG GCACCTTCAG GAATGGGACT TCCGGGTGAT GGATCTCCAA  
 10 401 GAGCGTCAAC GCCATCTTGC CTAGAGGAAC TTCAAGCAGA GATACAGGCA  
 451 GTTACTCAAG CTATCGATCA GATGTCAGAT GATTGA

The PSORT algorithm predicts inner membrane (0.4864).

The following *C.pneumoniae* protein (PID 4376431) was also expressed <SEQ ID 309; cp6431>:

1 LRAGGSLVTT YPKQGRLRS PEQLRVLDL VQSYPNHLHA IELDCGAIPQ  
 51 DLIGATYIIT FADFSTYILS LRSYQANSPP DDTWGIWFGS IDDPVQAVIS  
 15 101 FLKDHGFALP STLAQDPLLC TNK\*

The cp6431 nucleotide sequence <SEQ ID 310> is:

1 TTGCGAGCAG GAGGTAGTCT TGTTACAACA TACCCTAAGG AAGGTCAGAG  
 51 ATTGCGCTCC CCAGAACAGT TAAGAGTTCT GGATGATTTA GTGCAAAGCT  
 101 ATCCAAATCA CCTACATGCG ATTGAAGTTG ATTGTGGTGC AATCCCTCAA  
 20 151 GATTTGATCG GAGCCACCTA TATCATCACG TTCGCCGATT TTTCCACCTA  
 201 TATTCTCTCT TTAAGAAGCT ACCAAGCCAA TTCTCCCTCC GATGATACAT  
 251 GGGGGATTG GTTTGGATCT ATTGACGATC CTGTTCAAGC AGTCATATCA  
 301 TTTTAAAAG ATCATGATT TGCTCTTCCC TCGACCTTAG CTCAAGATCC  
 351 TTTGCTTTGT ACTACAAGT AA

25 The PSORT algorithm predicts cytoplasm (0.2115).

The following *C.pneumoniae* protein (PID 4376443) was also expressed <SEQ ID 311; cp6443>:

1 MIMTTISNSP SPALNPELSL IPPPTLVSSG TQTSLAYTIP AQGRRSTLRI  
 51 ILDIFIILG LATIISTFIV IFFLNLNLL STPSIISSSC LIIVGLLFLI  
 101 MGLYFMISSL DQGLVGLLQK ELSQAEEREE EYIQEIEALR GAPRAESPT  
 30 151 SPSTWL\*

The cp6443 nucleotide sequence <SEQ ID 312> is:

1 ATGATTATGA CTACTATATC TAACTCACCC TCCCCTGCAT TGAATCCCGA  
 51 ACTTTCCCTT ATTCCTCCAC CAACACTTGT ATCTTCAGGT ACGCAAACAT  
 101 CTCTAGCTTA TACGATCCCC GCACAAGGAC GAAGATCCAC CCTACGTATT  
 35 151 ATATTAGATA TATTCAATTAT CATCTTGGT TTAGCTACGA TCATTCTTAC  
 201 CTTTATTGTT ATTTTCTTTT TAAATGGGCT GAACTTGCTC TCGACCCCAT  
 251 CTATTATCTC TTCGTATGTT TTAATCATTG TTGGATTGCT TTTTATTGATT  
 301 ATGGGGTTAT ATTTCAATGAT CTCGAGTTTG GATCAGGGGC TTGTAGGCCCT  
 351 TCTGCAAAAG GAACTCTCTC AAGCCGAAGA AAGAGAAGAA GAGTATATCC  
 40 401 AGGAAATCGA AGCTTTAAGA GGAGCTCCTA GAGCAGAATC TCCCACAGAG  
 451 TCTCCTAGTA CCTGGTTATG A

The PSORT algorithm predicts inner membrane (0.5585).

The following *C.pneumoniae* protein (PID 4376496) was also expressed <SEQ ID 313; cp6496>:

1 MLIGRYSSDD QFTEATKNTP TIILKGFVRD NLEGLTNPIS EIVSETSSSI  
 51 KDSVLRSLPI LGSILGCARL YSTLSTNDPL DETQEKIWHI IFGALETGLL  
 45 101 GILILFPKII FVILHCIFHL VIGFCK\*

The cp6496 nucleotide sequence <SEQ ID 314> is:

1 ATGCTAATAG GCAGATACAG TAGTGATGAC CAATTCACTG AAGCAACAAA  
 51 AAACACCCCA ACCATAATTA AGCTAGGTTT TGTTAGAGAT AATCTCGAGG  
 101 GATTAACGAA CCCTATCTCT GAAATCGTCT CGGAAACCTC CTCTTCTATT  
 50 151 AAAGATTCCG TTCTTCGCTC TCTTCCTATT TTAGGGTCCA TTTTAGGATG  
 201 CGCCCGACTT TACAGCACAC TCTCTACAAA TGATCCTCTT GACGAAACTC  
 251 AAGAAAAGAT TTGGCACACT ATATTGGAG CCTTAGAAAC CTTAGGCTTA  
 301 GGGATTCTCA TCCTCTTATT TAAAATTATT TTTGTTATAT TACACTGCAT  
 55 351 ATTTCACTA GTTATTGGGT TCTGCAAATA A

-174-

1 MRPHRKHVSS KSLALKQSAS THVEITTKAF RLSMPLKQLI LEKSDHLPPM  
 51 ETIRVVLTSK KDKLGTEVHV VASHGKEILQ TKVHNANPYT AVINAFKKIR  
 101 TMANKHSNKR KDRTHKDLGL AAKEERIAIQ EEQEDRLSNE WLPVEGLDAW  
 151 DSLKTLGYVP ASAKKKISKK KMSIRMLSQD EAIRQLESAA ENFLIFLNEQ  
 201 EHKIQCIIYK HDGNYVLIEP SLKPGFCI\*

The cp6881 nucleotide sequence <SEQ ID 326> is:

1 ATGAGACCTC ATCGTAAACA CGTATCATCT AAAAGCTTAG CTTTAAAGCA  
 51 ATCTGCATCA ACTCATGTAG AGATCACAAAC AAAAGCCTTT CGTCTCTCTA  
 101 TGCCTCTAAA ACAGCTGATC CTAGAGAAAA GCGACCACCT CCCCCCTATG  
 151 GAAACAATCC GTGTGGTGCT AACCTCTCAT AAAGATAAGC TAGGCACCGA  
 201 GGTGCATGTT GTAGCTTCTC ATGGCAAAGA AATCCTTCAA ACTAAGGTTC  
 251 ATAACGCAAA CCCATACACT GCAGTGATCA ATGCTTTTAA GAAAATCCGC  
 301 ACCATGGCAA ATAAGCACTC CAATAAACGT AAAGACAGGA CAAAACATGA  
 351 TCTAGGTCTT GCAGCAAAAG AAGAACGTAT CGCAATACAG GAAGAACAAG  
 401 AAGATCGCCT TAGCAACGAG TGGCTTCTTG TCGAAGGCCT CGATGCCTGG  
 451 GATTCTCTAA AAACCTTTGG GTATGTTCCC GCATCAGCGA AAAAGAAGAT  
 501 CTCCAAGAAA AAGATGAGCA TTCGTATGCT ATCTCAAGAC GAGGCTATCC  
 551 GCCAGCTAGA GTCTGCCGCA GAAAACCTCC TGATCTTCTT GAACGAGCAA  
 601 GAGCATAAAA TCCAATGCAT TTATAAAAAA CATGACGGCA ACTATGTCCT  
 651 TATTGAACCT TCCCTCAAGC CAGGATTCTG CATCTGA

The PSORT algorithm predicts cytoplasm (0.249).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 161A; 6441 = lanes 7-9; 6748 = lanes 2-3; 6881 = lanes 4-6). The recombinant protein was used to immunise mice, whose sera were used in Western blots (Figures 161B, 162 & 163) and for FACS analysis.

These experiments show that cp6441, cp6748 & cp6881 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequence alone.

#### Example 164 and

#### Example 165

#### Example 166

The following *C.pneumoniae* protein (PID 4376444) was expressed <SEQ ID 327; cp6444>:

1 MEQPNCVIQD TTTVLYALNS FDPRLSDDTH RLKQSPLEA ENALGEFIEG  
 51 LDTNSFPLEE VAIPILPGYH PKFYLSTFIDR DDQGVHYEVL DGVFLKTVAA  
 101 CIIENSFLTD SMSPELLSEV KEALKR\*

The cp6444 nucleotide sequence <SEQ ID 328> is:

1 ATGGAGCAAC CCAATTGTGT GATTCAGGAT ACTACAACCTG TTTTGTATGC  
 51 CTTAAATAGC TTTGATCCTA GACTTAGTGA TGACACTCAC AGACTTGGGA  
 101 AGCAATCACC TCTTGAAGCA GAAAATGCTC TTGGAGAATT TATTGAAGGT  
 151 TTGGATACAA ATAGCTTTCC TTTAGAGGAA GTTGCCATTC CCATCCTGCC  
 201 AGGTTATCAC CCTAAGTTTT ATTTATCTTT CATAGATAGG GACGATCAAG  
 251 GTGTCCACTA TGAAGTTTTA GATGGCGTAT TTTTAAAGAC AGTCGCTGCT  
 301 TGTATTATAG AGAACTCCTT CTTAACTGAT TCTATGAGCC CGGAGCTTCT  
 351 CAGCGAAGTT AAGGAAGCTC TGAAACGATG A

The PSORT algorithm predicts cytoplasm (0.2031).

The following *C.pneumoniae* protein (PID 4376413) was also expressed <SEQ ID 329; cp6413>:

1 MAVQSIKEAV TSAATSVGCV NCSREAIPAF NTEERATSIA RSVIAAIIAV  
 51 VAISLLGLGL VVLAGCCPLG MAAGAITMLL GVALLAWAIL ITLRLNIPK  
 101 ABIPSPGNNG EPNERNSATP PLEGGVAGEA GRGGGSPLTQ LDLNSGAGS\*

The cp6413 nucleotide sequence <SEQ ID 330> is:

1 ATGGCTGTTC AATCTATAAA AGAAGCCGTA ACATCAGCCG CAACATCAGT

The PSORT algorithm predicts periplasmic space (0.4044).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 159A; 6435 = lanes 2-4; 6477 = lanes 5-7). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 159B & 160) and for FACS analysis.

- 5 These experiments show that cp6477 & cp6435 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequences alone.

**Example 161 and**

**Example 162 and**

**Example 163**

- 10 The following *C.pneumoniae* protein (PID 4376441) was expressed <SEQ ID 321; cp6441>:

```

1  VEAGANVLVI DTAHAHSGKV FQTVLEIKSQ FPQISLVVGN LVTAEAAVSL
51  AEIGVDAVKV GIGPGSICTT RIVSGVGYPQ ITAITNVAKA LKNSAVTVIA
101 DGRIRYSGDV VKALAAGADC VMLGSLLAGT DEAPGDIVSI DEKLFKRYRG
151 MGSLGAMKQG SADRYFQTQG QKKLVPGGVE GLVAYKGSVH DVLYQILGGI
15  201 RSGMGYVGAE TLKDLKTKAS FVRITESGRA ESHIHNIYKV QPTLNY

```

The cp6441 nucleotide sequence <SEQ ID 322> is:

```

1  GTGGAAGCTG GAGCAAATGT TCTAGTCATT GACACAGCTC ATGCACACTC
51  TAAAGGAGTA TTCCAAACAG TTTTAGAAAT AAAATCCCAG TTCCACAAA
20  101 TTTCTTTAGT TGTAGGGAAT CTTGTTACAG CTGAAGCCGC AGTTTCCTTA
151 GCTGAGATTG GAGTTGACGC TGTAAAGGTA GGTATTGGCC CAGGATCTAT
201 CTGTACAACT AGAATCGTTT CAGGGGTCGG TTATCCACAA ATTACTGCCA
251 TTACAAACGT AGCAAAAGCT CTTAAAAACT CTGCCGTGAC TGTAAATTGCT
301 GATGGGAGAA TCCGCTATTC TGGAGATGTG GTAAAAGCAT TAGCAGCAGG
351 AGCAGACTGT GTCATGCTAG GAAGTTTGCT TGCAGGGACT GATGAAGCTC
25  401 CTGGGGATAT CGTTTCTATC GATGAGAAGC TTTTAAAG GTACCGCGGC
451 ATGGGATCTT TAGGCGCTAT GAAACAAGGA AGTGCTGACC GGTATTTTCA
501 AACACAGGGA CAGAAAAGC TGGTTCCTGG GGGAGTTGAA GGAAGTAGTCG
551 CTTATAAAGG CTCTGTCCAC GATGTCTCT ATCAAAATTT AGGAGGAATA
601 CGCTCAGGTA TGGGGTATGT TGGAGCTGAA ACTCTCAAAG ATTTAAAAAC
30  651 TAAGGCTTCC TTTGTTTCGAA TTAAGTGAATC TGAAGAGCT GAAAGTCATA
701 TTCATAATAT TTACAAAGTT CAACCAACCT TAAATTATTA A

```

The PSORT algorithm predicts bacterial inner membrane (0.132).

The following *C.pneumoniae* protein (PID 4376748) was also expressed <SEQ ID 323; cp6748>:

```

1  LFSEGTALNL FRIFAPLRNR VTTEYSRARQ PDLHRIAIVY IGVLDSSESSK
35  51  ILERLISYMS CIYSESQMYL RFFMGKVNQ SAVLSKLHVE NLHIRCGFFS
101 EDAVPESEPF DLSIVVHTDR SCPLPTKKRS SSWELQTVEL PESIYPQSEF
151 LLMRPRMLS*

```

The cp6748 nucleotide sequence <SEQ ID 324> is:

```

1  TTGTTCTCTG AGGGGACAGC TCTAAATTTA TTTCGTATAT TTGCTCCACT
40  51  ACGCAACCGT GTGACTACAG AATACAGTCG TGCTAGGCAA CCCGACCTAC
101 ATAGAATTGC CATCGTCTAT ATAGGAGTTC TCGATTGAGA AAGTTCCAAG
151 ATCCTAGAGC GGCTAATCTC TTATATGAGT TGTATCTATT CTGAATCGCA
201 AATGTATTTA AGATTCTTTA TGGGCAAGAA TGTAAATCAA AGTGCTGTAC
251 TCTCAAAATT ACATGTAGAA AATCTGCACA TCCGTTGTGG GTTTTTCAGC
45  301 GAGGATGCTG TTCCAGAGAG TGAGCCCTTC GATCTCTCCA TCTACGTGCA
351 CACAGATCGT AGCTGTCTTC TCCCTACGAA AAAACGGAGC AGCTCCTGGG
401 AACTCCAAAC TGTAAGACTC CCAGAGTCAA TATATCCACA GTCGGAATTC
451 CTATTGATGA GACCTCGAAT GCTTTCGTAG

```

The PSORT algorithm predicts cytoplasm (0.170).

- 50 The following *C.pneumoniae* protein (PID 4376881) was also expressed <SEQ ID 325; cp6881>:

-176-

201 AACAGAGAAG ACCACGACCC GTCATTTGGT GCTCTCTATT CGCCATAACG  
 251 CCTCTCTTAT TGTAATTCGT ACGGTTCTTG GTTCAGCTTC TTGGATCGCT  
 301 GCTTTGTTAG ATCAAGGGCT CAAAGATGAA ATTCTTGGAA CTTTGGCAGG  
 351 AGATGACACG ATTTTGTGCA CTCCTATAGA TGAAGGGAGG CTCCCATTGT  
 401 TGATGGTTTC GATTGCAAAT TTAAGGCAAG TTTTCTTGA TTA

The PSORT algorithm predicts inner membrane (0.1510).

The following *C.pneumoniae* protein (PID 4376540) was also expressed <SEQ ID 335; cp6540>:

1 MSQCQSSSTS TWEWMKSFVP NWKNPTPLS PIPSEDEFIL AYEPFVLPKT  
 51 DPENANQANPP GTSTPNVENG IDDLNPLLGQ PNEQNNANNP GTSGSNPTSL  
 101 PAPERLPETE ENSQEEEQGS QNNEDLIG\*

The cp6540 nucleotide sequence <SEQ ID 336> is:

1 ATGTCTCAAT GTCAGAGTAG CAGTACATCT ACCTGGGAAT GGATGAAATC  
 51 TTTTGTGCCA AACTGGAAGA ATCCAACCTC CCCCTTATCT CCTATACCTT  
 101 CTGAGGACGA ATTTATATTA GCATACGAGC CATTGTGTCT ACCGAAAACA  
 151 GATCCAGAAA ACGCACAAAGC TAATCCTCCA GGCACATCTA CACCGAATGT  
 201 AGAAAACGGG ATCGATGATC TCAACCTCT TCTGGGGCAA CCCAACGAAC  
 251 AAAACAATGC CAACAATCCA GGAACCTCTG GATCTAATCC TACATCTCTA  
 301 CCCGCCCCCG AACGACTCCC TGAACTGAA GAGAACAGCC AAGAAGAAGA  
 351 ACAAGGATCT CAAAATAATG AGGATCTTAT AGGATAA

The PSORT algorithm predicts cytoplasm (0.3086).

The following *C.pneumoniae* protein (PID 4376743) was also expressed <SEQ ID 337; cp6743>:

1 LREEGSVSFR EYFRAYMCDK IVAQKNFLFT LDAVIKQAGW RSQEKLNLFY  
 51 VESQALGREI KVSLEEIYQS MVGILGSQRT KKSFKFSVDF TPLEQALQER  
 101 CSSDDDEDAT ATSTATGATA SPTDMHEDE\*

The cp6743 nucleotide sequence <SEQ ID 338> is:

1 TTGAGAGAAG AAGGTAGTGT TTCTTTCAGA GAATATTTCA GAGCCTATAT  
 51 GTGTGATAAA ATCGTGGCAC AGAAGAAGCTT CTTATTTACT TTAGACGCTG  
 101 TAATTAAACA GGCCGGTTGG AGATCACAAG AGAACTCAA TTTATTTTAT  
 151 GTTGAAAGTC AGGCTTTAGG AAGAGAAATC AAAGTCAGCT TAGAGGAATA  
 201 TATTGAGAGT ATGGTCGGGA TTTTGGGATC TCAGAGAAC AAGAAAAGCT  
 251 TTAAGTTTTC TGTGACTTTT ACCCTTTTAG AGCAGGCTCT ACAAGAAAGA  
 301 TGCTCTTCTG ATGATGACGA AGATGCAACA GCAACTTCGA CCGCTACAGG  
 351 GGCAACAGCA TCTCCGACTG ACATGCACGA AGATGAGTAA

The PSORT algorithm predicts cytoplasm (0.2769).

The following *C.pneumoniae* protein (PID 4377041) was also expressed <SEQ ID 339; cp7041>:

1 MLMMLMMIIG ITGSGGAGKT TLTQNIKEIF GEDVSVICQD NYYKDRSHYT  
 51 PEERANLIWD HPDAFDNDLL ISDIKRLKNN EIVQAPVDF VLGNRSKTEI  
 101 ETIYPSKVL VEGILVFENQ ELRLMDIRI FVDTDADERI LRRMVRDVQE  
 151 QGDSVDCIMS RYLSMVKPMH EKFIETPKY ADIIVHGNR QNVVTNLSQ  
 201 KIKNHLENAL ESDETYVMVN SK\*

The cp7041 nucleotide sequence <SEQ ID 340> is:

1 ATGTTGATGA TGCTTATGAT GATTATTGGA ATTACAGGAG GTTCTGGAGC  
 51 TGGGAAAACC ACCCTAACCC AAAACATTAA AGAAATTTTC GGTGAGGATG  
 101 TGAGTGTTAT CTGCCAAGAT AATTATTACA AAGATAGATC TCATTATACT  
 151 CCTGAAGAAC GTGCCAATTT AATTGGGAT CATCCGGACG CCTTTGATAA  
 201 TGAATTATTA ATTTGAGACA TAAAACGTCT AAAAAATAAT GAGATTGTCC  
 251 AAGCCCCAGT TTTTGATTCT GTTTTAGGTA ATCGATCTAA AACGGAGATA  
 301 GAAACGATCT ATCCATCTAA AGTTATTCTT GTTGAAGGTA TTCTGGTCTT  
 351 TGAAAATCAA GAAGTAGAG ATCTTATGGA TATTAGGATC TTTGTAGACA  
 401 CCGATGCTGA TGAAAGGATA CTACGCCGTA TGGTTCGAGA TGTTCAGAA  
 451 CAAGGAGATA GCGTGGACTG CATCATGTCT CGTTATCTTT CTATGGTAAA  
 501 GCCTATGCAT GAGAAATTTA TAGAGCCGAC TCGGAAATAT GCTGATATCA  
 551 TTGTACATGG AAATTACCGA CAAAACGTAG TAACAAATAT TTTGTACACG  
 601 AAAATTAAAA ATCATTTAGA GAATGCCCTG GAAAGCGATG AGACGTATTA  
 651 TATGGTCAAC TCTAAGTAA

-175-

51 AGGATGTGTA AACTGTTCTA GAGAGGCTAT ACCAGCATTT AATACAGAGG  
 101 AGAGAGCAAC GAGTATTGCT AGATCTGTTA TAGCAGCTAT CATTGCTGTT  
 151 GTAGCTATCT CCTACTCGG ACTAGGTCTT GTAGTTCTTG CTGGTTGCTG  
 201 TCCTTTAGGA ATGGCTGCGG GTGCTATAAC AATGCTGCTG GGTGTAGCAT  
 251 TATTAGCTTG GGCAATACTG ATTACTTTGA GACTGCTTAA TATACCTAAG  
 301 GCTGAAATAC CGAGTCCAGG GAACAACGGT GAGCCTAATG AAAGAAATTC  
 351 AGCAACTCCT CCTCTAGAGG GTGGTGTTCG AGGAGAAGCC GGTGCGGGCG  
 401 GGGGGTCACC TTAAACCCAA CTTGATCTCA ATTCAGGGGC GGAAGTTAG

The PSORT algorithm predicts inner membrane (0.6180).

10 The following *C.pneumoniae* protein (PID 4377391) was also expressed <SEQ ID 331; cp7391>:

1 MMLRVIELPL LPIKQALEKA FVQYNSYKAK LTKVEPCFRE SPAYITSEER  
 51 LQSLDQTLER AYKEYQKRFO EPSRLESEVS GCREHLREQV KQFETQGLDL  
 101 IKEELIFVSD VLFKRMVSCL VSTVHVFFME FYYEYFELHR LRLRAQWMAN  
 151 AEIYSKVRKA FPEMLKETLE KAKAPREEEY WLLCEERKSK EKRLILNKIE  
 201 AAQQRVKDLE PPIKETGKQ KKKKEYSFFI RLKS\*

The cp7391 nucleotide sequence <SEQ ID 332> is:

1 ATGATGCTTC GTGTCATAGA GCTTCCACTA CTTCTATAA AGCAAGCGTT  
 51 GGAGAAGGCT TTTGTACAAT ATAATAGCTA CAAAGCGAAG TTAACCAAGG  
 101 TAGAACCTTG CTTTAGAGAG AGCCCTGCCT ATATAACTAG CGAAGAGCGA  
 151 CTCCAGAGTT TGGATCAGAC TTTAGAACGT GCGTACAAAG AGTACCAGAA  
 201 GAGATTCCAG GAGCCTTCAC GTTTGGAATC GGAAGTAAGT GGATGTAGAG  
 251 AGCATCTTAG AGAGCAGGTA AAACAATTTG AAACCTCAAGG ACTAGACTTG  
 301 ATCAAAGAAG AGCTTATTTT TGTTAGTGAT GTGTATTCC GAAAAATGGT  
 351 CAGTTGTCTA GTGTCGACAG TGCATGTTCC CTTTATGGAG TTTTATTATG  
 401 AGTATTTTGA GTTGCATAGA TTGAGGTTGC GGGCCCAATG GATGGCGAAT  
 451 GCCGAGATTT ATAGCAAAGT TAGAAAAGCA TTCCAGAGA TGTGAAGGA  
 501 GACCTTAGAA AAAGCTAAGG CTCCAGAGA AGAAGAGTAT TGGTTACTTT  
 551 GCGAGGAGAG AAAGAGTAAG GAGAAGCGTT TGATTTCTCA CAAGATAGAG  
 601 GCAGCTCAGC AGCGGGTAAA AGATTTAGAA CCTCCTCCTA TTAAAGAGAC  
 651 AGGGAACAG AACCGGAAGA AAGAATATTC GTTTTTCATT CGATTAAAAAT  
 701 CGTGA

The PSORT algorithm predicts inner membrane (0.1489).

The proteins were expressed in *E.coli* and purified as his-tag and GST-fusion products (Figure 164A; 6444=lanes 11-12; 7391=lanes 2-3; 6413=lanes 4-6). The recombinant protein was used to immunise mice, whose sera were used in Western blots (Figures 164B, 165 & 166) and for FACS analysis.

These experiments show that cp6444, cp6413 & cp7391 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequence alone.

Example 167 ,  
 Example 168 ,  
 Example 169 and  
 Example 170

The following *C.pneumoniae* protein (PID 4376463) was expressed <SEQ ID 333; cp6463>:

1 MKKKVTIDEA LKEILRLEGA ATQEELCAKL LAQGFATTQS SVSRWLRLKIQ  
 51 AVKVAGERGA RYSLPSSTK TTRHLVLSI RHNASLIVIR TVPGSASWIA  
 101 ALLDQGLKDE ILGTLAGDDT IFVTPIDEGR LPLLMVSIAN LLQVFLD\*

The cp6463 nucleotide sequence <SEQ ID 334> is:

1 ATGAAAAAAA AAGTAACTAT AGATGAGGCT TTAAAAGAAA TTTTACGTCT  
 51 TGAAGGAGCG GCAACTCAGG AGGAATTATG TGCAAACTC TTAGCTCAAG  
 101 GTTTTGCTAC AACCCAGTCG TCTGTATCTC GTTGGCTACG AAAGATTACG  
 151 GCTGTAAAGG TTGCTGGAGA GCGTGGTGCT CGTTATTCTT TACCCTCTTC

201 TTGGTATACA AGTGACGAAG ATTGGAAAAA ACAAGTGGTT TGA

The PSORT algorithm predicts inner membrane (0.145).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 171A; 6632 = lanes 5-7; 6648 = lanes 8-10; 6497 = lanes 2-4). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 171B, 172, 173) and for FACS analysis.

These experiments show that cp6632, cp6648 and cp6497 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequence alone.

Example 174 ,  
Example 175 ,  
Example 176 ,  
Example 177 and  
Example 178

The following *C.pneumoniae* protein (PID 4377200) was expressed <SEQ ID 347; cp7200>:

1 MPVPIDNSSR NLQEVPESE DLEQHAESP THQSAESSSL QLSLASSAIS  
51 SRVEQLSSLV LGMENSDFSS LRDVPIFSAI YESSTHTPVP TPLVGVGYIN  
101 GSQSGYYDTQ RESLHLSQLL GSRRVEVVYN QGNFMEASLL NLCPRRPRRD  
151 PSPISLALLE LWEAFFLEHP PGSTFNPIFF W\*

The cp7200 nucleotide sequence <SEQ ID 348> is:

1 ATGCCCGTTC CTATAGATAA TTCCTCTCGC AACCTACAAG AAGTTCCAGA  
51 AAGCCTAGAA GACCTCGAAC AACACGCAGA AGAATCTCCT ACTCATCAAA  
101 GTGCAGAAAG CAGTTCTTTG CAACTGTCTC TAGCCTCCTC AGCAATTCTT  
151 AGTAGAGTAG AACAACATATC TTCCTCTCGT TTAGGAATGG AAAATTCAGA  
201 TTTCTCCTCT TTAAGAGACG TTCCTATCTT CTCAGCTATC TACGAATCTT  
251 CAACACACAC ACCTGTCCCC ACTCCTCTAG TTGGCGTGGG ATATATCAAC  
301 GGAAGTCAAT CAGGATACTA CGATACACAA AGAGAATCTC TTCACCTCAG  
351 CCAATTGTTA GGAAGCCGAA GAGTTGAAGT TGTCTATAAC CAAGGAAACT  
401 TCATGGAGGC CTCTTTGCTA AATCTGTGCC CCAGAAGACC TCGAAGAGAT  
451 CCCTCTCCAA TTTCTTTAGC TCTATTAGAG CTCTGGGAAG CATTTTTTTT  
501 AGAACACCCC CCAGGTAGCA CTTTTAATCC AATATTTTTT TGGTAA

The PSORT algorithm predicts cytoplasm (0.3672).

The following *C.pneumoniae* protein (PID 4377235) was also expressed <SEQ ID 349; cp7235>:

1 LNFVSTLTGS DFYAPVLEKL EEFADTTGQ VILFSSSPDF IVHPIAQQLG  
51 ISSWYASCYR DQSAEQTIYK KCLTGDKKAQ ILSYIKKINQ ARSHTFSDHI  
101 LDLPFLMLGE EKTVVRPQGR LKKMAKKYYW NIV\*

The cp7235 nucleotide sequence <SEQ ID 350> is:

1 TTGAATTTTG TATCGACTCT GACCGGCTCC GATTTTATG CTCCTGTTT  
51 AGAAAACTA GAAGAAGCTT TTGCAGATAC CACAGGACAG GTGATCCTTT  
101 TTTCTTCTTC TCCAGACTTT ATTGTCCACC CCATAGCGCA GCAACTCGGG  
151 ATTAGTTCTT GGTATGCGTC GTGTTATCGC GATCAGTCTG CAGAACAGAC  
201 GATCTATAAA AAATGTCTTA CAGGGGATAA AAAAGCGCAA ATTTTGAGTT  
251 ATATTAAAAA AATTAATCAA GCAAGAAGCC ATACCTTCTC CGACCATATT  
301 TTAGATCTTC CTTTCTTAT GCTGGGAGAA GAGAAAACCG TCGTTCGCCC  
351 TCAGGGACGA CTCAAGAAAA TGGCAAAAAA ATATTACTGG AATATCGTTT  
401 AA

The PSORT algorithm predicts cytoplasm (0.3214).

The following *C.pneumoniae* protein (PID 4377268) was also expressed <SEQ ID 351; cp7268>:

1 MMHRYFIPLL ALLIFSPSLV RAELOPSENK KGGWPTQLSC AEGSQLFCKF



The PSORT algorithm predicts inner membrane (0.1022).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 167A; 6463 = lanes 2-4; 6540 = lanes 5-7; 6743 = lanes 8-9; 7041 = lanes 10-11). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 167B, 168, 169 & 170) and for FACS analysis.

These experiments show that cp6463, cp6540, cp6743 & cp7041 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequence alone.

#### Example 171 and Example 172 and Example 173

The following *C.pneumoniae* protein (PID 4376632) was expressed <SEQ ID 341; cp6632>:

```

1  VQLFQYMNES GWDWLCDDFS QGEGFQLSRL VGLLHSSWAL YEAKEQFYLP
51  EVSLLTWEEEL IEMQLLSKPT KHGVAKDLN VFEKHFQRF R QYLGSLDLNQ
101 RFENTFLNYP KYHLDRE*

```

The cp6632 nucleotide sequence <SEQ ID 342> is:

```

1  GTGCAATTAT TTCAATATAT GAATGAGTCC GGATGGGATT GGCTTTGTGA
51  TTTTGATTCT CAAGGCGAGG GATTCCAGTT ATCACGTCTG GTTGGGCTGT
101 TACATTCGTC CTGGGCATTA TACGAAGCAA AAGAGCAATT TTACCTTCCT
151 GAGGTTTCTC TATTGACCTG GGAAGAACTG ATAGAAATGC AGTTATTAAAG
201 CAAACCAACA AAACACGGGG TTGCAAAGA TCTTTGTAAT GTATTTGAAA
251 AACACTTTCA AAGGTTTAGA CAGTACCTAG GTTCCTTAGA TCTAAATCAA
301 AGGTTTCGAAA ATACCTTCTT GAATTATCCT AAATACCAT TTAGATAGGGA
351 GTGA

```

The PSORT algorithm predicts cytoplasm (0.3627).

The following *C.pneumoniae* protein (PID 4376648) was also expressed <SEQ ID 343; cp6648>:

```

1  MPVSSAPLPT SHRPSSGNLG LMEPNKALK AKHQDKTKT IKLLVKILVA
51  ILVIEVLGII AAFIPGTPP ICLIIIGGLI LTTVLCVLLL VIKLALVNKT
101 EGTAEQQIK RKLSSKSIS*

```

The cp6648 nucleotide sequence <SEQ ID 344> is:

```

1  ATGCCCGTGT CCTCAGCCCC CCTACCCACA AGCCACCGCC CTTCTCTGCG
51  AAATCTAGGC CTCATGGAAC CAAATCCCAA AGCTCTAAAA GCAAAGCATC
101 AAGATAAAAC GACGAAGACG ATTAACTTT TAGTTAAAT CTTGTGTGCC
151 ATTCTAGTAA TAGAAGTTT AGGAATAATT GCAGCTTTCT TTATTCCTGG
201 GACTCCTCCC ATCTGCTTGA TTATCCTAGG AGGCCTTATT CTTACAACAG
251 TACTCTGTGT GCTTCTTCTT GTTATAAAGC TTGCCCTTGT AAACAAAACC
301 GAAGGAACAA CTGCTGAACA GCAGATAAAA CGTAACTCT CTTCTAAAAG
351 TATTTCTTAG

```

The PSORT algorithm predicts inner membrane (0.6074).

The following *C.pneumoniae* protein (PID 4376497) was also expressed <SEQ ID 345; cp6497>:

```

1  MKPNSIIFLE NTKHYPDIFR EGFVRDRHGL MEASDWLLST EITIIRSILG
51  AIPILGNILG AGRLYSVWYT SDEDWKKQVV *

```

The cp6497 nucleotide sequence <SEQ ID 346> is:

```

1  ATGAAGCCAA ATAGTATTAT TTTTGTAGAA AATACTAAGC ATTATCCCCGA
51  CATCTTTTCA GAAGGATTGT TTCGTGATCG TCATGGACTA ATGGAAGCCT
101 CGGATTGGTT ACTTTCTACG GAAATTACGA TCATTGCTC CATTCTGGGA
151 GCTATCCCTA TTTTAGGAAA TATTCTTGGA GCCGGACGAC TCTATAGCGT

```

-180-

101 TAGTTTTTGG TATGCTCTTA CTGATTTTCAG GAGCTCTCTT TCTGACGTTA  
 151 GGGATTCCAG GATTGAGTGC AGCAATTTCT TTTGGATTAG GCATCCGTCT  
 201 CTCCGCATTA GGAGGAGTGC TGATGATTTT GGGACTACTA TGTCTTTTAG  
 251 TAAAACGAGA GATTCCGACA GTACGACCAG AAGAAATTC TGAAGGGGTT  
 5 301 TCGCTGGCTC CTTCTGAGGA GCCAGCTCTA CAGGCAGCTC AGAAGACTTT  
 351 AGCTCAGCTG CCTAAGGAAT TGGATCAGTT AGATACAGAT ATTCAGGAAG  
 401 TGTTCGCATG TTTAAGAAAG CTGAAAGATT CTAAGTATGA AAGTCGAAGT  
 451 TTTTAAACG ATGCTAAGAA GGAGCTTCGA GTTTTTGACT TTGTGGTTGA  
 10 501 GGATACCTC TCGGAGATTT TCGAGTTGCG GCAGATTGTG GCTCAAGAGG  
 551 GATGGGATTT AAACCTTTTG ATCAATGGGG GACGAAGCCT CATGATGACT  
 601 GCAGAATCTG AATCGCTTGA TTTGTTTCAT GTATCGAAGC GGCTAGGGTA  
 651 TTTACCTTCT GGGGATGTTT GAGGGGAGGG GTTAAAGAAA TCTGCGAAGG  
 701 AGATAGTCGC TCGTTTGATG AGCTTGCAAT GCGAGATTCA CAAGGTGGCG  
 751 GTAGCGTTTG ATAGGAATTC CTATGCGATG GCAGAAAAGG CGTTTGCGAA  
 15 801 AGCGTTGGGA GCTTTAGAAG AGAGTGTGTA TCGGAGTCTG ACGCAGAGTT  
 851 ATAGAGATAA ATTTTTGGAG AGCGAGAGGG CGAAGATCCC ATGGAATGGG  
 901 CATATAACCT GGTAAAGAGA TGATGCGAAG AGTGGGTGTG CTGAAAAGAA  
 951 GCTTCGGGAT GCCGAGGAAC GTTGAAGAA ATTTAGGAAA GCAGTCTTTT  
 1001 GGGTAGAAGA AGACGGGGGC TTTGACATCA ATAATCTCCT TGGAGACTGG  
 20 1051 GGGACAGTGC TTGATCCTTA TAGACAAGAG AGAATGGACC AGATAACGTT  
 1101 CCATGAGTTG TATGAAAAAA CTACGTTTTT GAAAAGACTG CACAGAAAGT  
 1151 GTGCGTTAGC GAAAACAACC TTTGAAAAGA AGAGATCTAA AAAGAAATTTG  
 1201 CAGGCAGTCG AGGAGGCGAA TGCACGTAGG TTGAAATATG TAAGGGATTG  
 1251 GTATGATCAG GAGTTTCAGA AAGCAGGGGA GAGATTAGAG AAACATGCATG  
 25 1301 CTTTGTATCC TGAGGTTTCA GTCTCTATAA GAGAGAACAA AATACAAGAG  
 1351 ACGCGCTCTA ATTTAGAGAA AGCCTATGAG GCTATCGAAG AGAACTATCG  
 1401 TTGCTGTGTC CGAGAGCAAG AGGACTACTG GAAAGAAGAA GAGAAAAGGG  
 1451 AAGCGGAGTT TAGGGAGAGG GGAAACAAGA TTCTTTCTCC TGAGGAGCTG  
 30 1501 GAAAGTTCTT TGGAGCAATT CGACCATGGT TTGAAAAATT TTTCTGAGAA  
 1551 ATTAATGGAA TTGGAAGGGC ATATCTTAAA ACTTCAGAAA GAAGCCACAG  
 1601 CAGAGGTGGA GAATAAAATA CTTTCAGATG CAGAGAGCCG CCTTGAGATT  
 1651 GTATTTGAAG ATGTCAAGGA GATGCCCTGT CGAATTGAGG AGATAGAGAA  
 1701 GACGCTGCGT ATGGCGGAGC TGCCCCTACT TCCTACGAAG AAGGCGTTTG  
 1751 AGAAGGCCCTG CTCACAATAT AATAGCTGCG CAGAGATGTT GGAGAAGGTG  
 35 1801 AAGCCTTACT GCAAGGAGAG CCTCGCCTAT GTGACTAGCA AAGAGCGTTT  
 1851 AGTGAGCTTG GATGAAGATT TACGACGAGC CTACACAGAG TGTCAGAAGA  
 1901 GATTCCAGGG GGATTCCGGT TTGGAGTCGG AAGTAAGAGC CTGTCCGAGAG  
 1951 CAACTGCGAG AGCGGATCCA AGAGTTTGAA ACTCAAGGGC TGGACTTGGT  
 2001 GGAAAAAGAG TTGCTTTGTG TGAGTAGTAG ATTAAGAAAT ACAGAGTGCG  
 40 2051 ATTGTTATC TGGTGTAAAG AAAGAAGCAC CTCCTGGTAA GAAGTTTAT  
 2101 GCCCAGTATT ATGATGAGAT TTATCGAGTT AGAGTTCAAT CCCGATGGAT  
 2151 GACGATGTCT GAGAGATTGA GAGAGGGAGT TCAAGCATGC AACAAGATGT  
 2201 TGAAGGCAGG CCTAAGCGAA GAAGATAAGG TTCTTAAAGA AGAAGAGTAT  
 2251 TGGTTGTATC GAGAGGAGAG AAAGAATAAA GAGAAACGTT TGGTTGGTAC  
 45 2301 TAAGTAGTA GCAACGCAGC AGCGAGTTGC AGCATTTGAA TCCATAGAAG  
 2351 TTCTGAGAT TCCTGAGGCC CCAGAGGAGA AACCGAGTTT GCTGGATAAA  
 2401 GCGCGTTCTT TATTTACTCG CGAGGACCAT ACCTAG

The PSORT algorithm predicts inner membrane (0.461).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 174: 7200=lanes 2-3;  
 50 7236=lanes 4-5; 7268=lanes 6-8; 7375=lanes 9-10; 7388=lanes 11-12). The recombinant proteins  
 were used to immunise mice, whose sera were used in Western blots (Figures 174, 175, 176, 177 &  
 178) and for FACS analysis.

These experiments show that cp7200, cp7235, cp7268, cp7375 & cp7388 are surface-exposed and  
 immunoaccessible proteins and that they are useful immunogens. These properties are not evident  
 55 from the sequence alone.

### Example 179

The following *C.pneumoniae* protein (PID 4376723) was expressed <SEQ ID 357; cp6723>:

-179-

51 EAAYNNAIEE GKPGILVFFS ERPTPEFADL TNGSFSLSTP IAKGFNVVVL  
 101 CPGLISPLDF FHKMDPVILY MGSFLEMPFE VEAVSGPRLC YILIDEQGGA  
 151 QCQAVLPLET KN\*

The cp7268 nucleotide sequence <SEQ ID 352> is:

5 1 ATGATGCACC GTTATTTTAT TCCTTTATTA GCACTTCTCA TTTTCTCTCC  
 51 TTCTTTAGTC AGGGCAGAGC TACAACCAAG TGAAAACAGA AAAGGGGGGT  
 101 GGCCATACACA ACTTTCCTGT GCAGAAGGTT CGCAACTCTT CTGTAAATTC  
 151 GAAGCTGCCT ATAATAATGC AATTGAGGAA GGGAAACCTG GGATTTTAGT  
 201 CTTTTTCTCT GAGCGACCCA CACCAGAATT TGCCGACTTA ACGAATGGTT  
 10 251 CATTTTCTCT CTCTACGCCA ATCGCCAAGG GCTTTAATGT CGTTGTGTTA  
 301 TGCCCCGGGC TTATCAGTCC CTTAGACTTT TTCCACAAA TGGATCCTGT  
 351 GATTCTCTAT ATGGGAAGTT TTCTAGAGAT GTTCCCTGAA GTGGAGGCAG  
 401 TTAGTGGCCC TCGCTTATGT TATATCTTAA TAGATGAACA GGGTGGGGCT  
 451 CAATGTCAGG CTGTCCTGCC TTTAGAAACA AAGAATTAG

15 The PSORT algorithm predicts inner membrane (0.1235).

The following *C.pneumoniae* protein (PID 4377375) was also expressed <SEQ ID 353; cp7375>:

1 MQRIIIIVGID TGVGKTIVSA ILARALNAEY WKPIQAGNLE NSDSNIVHEL  
 51 SGAYCHPEAY RLHKPLSPHK AAQIDNVSIE ESHICAPKTT SNLIIETSGG  
 101 FLSPCTSKRL QGDVFSSWSC SWILVSQAYL GSINHNTCLTV EAMRSRNLNI  
 20 151 LGMVVNGYPE DEEHWLTQEI KLPIIGTLAK EKEITKTIIS CYAEQWKEVW  
 201 TSNHQGIQGV SGTPSLNLH\*

The cp7375 nucleotide sequence <SEQ ID 354> is:

1 ATGCAACGTA TCATCATTGT AGGAATCGAC ACTGGCGTAG GAAAAACCAT  
 51 TGTCAAGTCT ATCCTTGCTA GAGCACTTAA CGCAGAATAC TGGAAACCTA  
 25 101 TACAAGCAGG GAATCTAGAA AATTCAGATA GCAATATTGT TCATGAGCTA  
 151 TCGGGAGCCT ACTGTCATCC CGAAGCTTAT CGATTGCATA AGCCCTTGTC  
 201 TCCACACAAG GCAGCGCAA TCGATAATGT AAGTATCGAA GAGAGTCATA  
 251 TTTGTGCCCC AAAAACAAC TCGAATCTGA TTATTGAGAC TTCAGGAGGA  
 30 301 TTTTATATCCC CCTGCACATC AAAAAGACTT CAGGGAGATG TGTTCCTTC  
 351 TTGGTCATGT TCTTGGATTT TAGTGAGCCA AGCATATCTC GGAAGTATCA  
 401 ATCACACCTG TTTAACGGTA GAAGCAATGC GCTCACGAAA CCTCAATATC  
 451 TTAGGTATGG TGGTAAATGG GTATCCAGAG GACGAAGAGC ACTGGCTAAC  
 501 TCAAGAAATC AAGCTTCCTA TAATCGGGAC TCTTGCCAAG GAAAAAGAAA  
 551 TCACAAAGAC AATCATAAGC TGTTATGCCG AACAATGGAA GGAAGTATGG  
 35 601 ACAAGCAATC ATCAGGGAAT TCAGGTGTA TCTGGCACCC CTTCACTCAA  
 651 TCTGCATTAG

The PSORT algorithm predicts cytoplasm (0.0049).

The following *C.pneumoniae* protein (PID 4377388) was also expressed <SEQ ID 355; cp7388>:

1 MQVLLSPQLP PPPQHSVGS I SSPSKLRVLA ITFLVFGMLL LISGALFRTL  
 40 51 GIPGLSAAIS FGLGIGLSAL GGVLMISGLL CLLVKREIPT VRPEEIPGV  
 101 SLAPSEEPAL QAAQKTLAQL PKELDQLDTD IQEVFACLRK LKDSKYESRS  
 151 FLNDAKKELR VFDFVVEDTL SEIFELRQIV AQEGWDLNPL INGGRLMMT  
 201 AESESLDLFH VSKRLGYLPS GDVRGEGLKK SAKETVARLM SLHCEIHKVA  
 251 VAFDRNSYAM AEKAFKALG ALEESVYRSL TQSYRDKFLE SERAKIPWNG  
 45 301 HITWLRDDAK SGCAEKKLRD AEERWKKFRK AVFWVEEDGG FDINNLLGDW  
 351 GTVLDPYRQE RMDEITFHEL YEKTTFKRL HRCALAKTT FEKKRSKKNL  
 401 QAVEEANARR LKYVRDWYDQ EFQKAGERLE KLHALYPEVS VSIRENKIQE  
 451 TRSNLEKAYE AIEENYRCCV REQEDYWKKE EKREAEFRER GNKILSPEEL  
 501 ESSLEQFDHG LKNFSEKLME LEGHILKLQK EATAEVENKI LSDAESRLEI  
 55 551 VFEDVKEMPC RIEEIEKTLR MAELPLLPTK KAFEKACSQY NSCAEMLEKV  
 601 KPYCKESLAY VTSKERLVSL DEDLRRAYTE CQKRFQGDSG LESEVRACRE  
 651 QLRERIQEFE TQGLDLVEKE LLCVSSRLRN TECDCVSGVK KEAPPGKKFY  
 701 AQYYDEIYRV RVQSRWMTMS ERLREGVQAC NKMLKAGLSE EDKVLKEEY  
 751 WLYREERKNK EKRLVGTKIV ATQQRVAAFE SIEVPEIPEA PEEKPSLLDK  
 55 801 ARSLFTREDH T

The cp7388 nucleotide sequence <SEQ ID 356> is:

1 ATGCAAGTAC TTCTATCTCC GCAGCTACCC CCCCCCCCCC AACACTCTGT  
 51 AGGGTCGATT TCTTCTCCAT CTAAACTTCG CGTTTTAGCG ATTACTTTTT

Example 181 ,  
 Example 182 ,  
 Example 183 ,  
 Example 184 and  
 Example 185

The following *C.pneumoniae* protein (PID 4376301) was expressed <SEQ ID 361; cp6301>:

```

1  LNQDLQNVYQ ECQKATGLES EVSAYRDHLR EQITEFETQG LDVIKEELLF
51  VSSTLKSLS YDPLIADIPC MKFYEEYDYG IDKARVQSRW LEKSERYRKA
101 KKGFOEMLKE GLFKEDQALK KAEYRLREK RMNKEKLLIC NKIEAAQQRV
151 QEFGPSDS*

```

The cp6301 nucleotide sequence <SEQ ID 362> is:

```

1  TTGAATCAGG ATTTACAAAA TGTATACCAA GAGTGCCAGA AGGCTACAGG
51  TTTAGAATCG GAAGTGAGTG CATATAGAGA TCATCTTAGA GAGCAGATCA
101 CAGAGTTTGA AACTCAAGGG CTGGACGTGA TAAAAGAAGA ACTTCTTTT
151 GTGAGTAGTA CTCTCAAAAG TAAATTGAGC TATGATCCAT TAATAGCAGA
201 CATTCCCTGT ATGAAGTTTT ATGAGGAGTA TTATGATGGC ATTGATAAAG
251 CGAGAGTTCA ATCCCGATGG CTGGAGAAGT CTGAGAGGTA TAGAAAGGCG
301 AAGAAGGGAT TCCAAGAGAT GCTGAAGGAA GGCCTATTCA AAGAAGATCA
351 GGCTTTGAAA AAAGCAGAGT ATAGATTACT TCGAGAGAAG AGAATGAATA
401 AGGAGAAGCT TTTGATTTCG AATAAGATAG AAGCAGCTCA GCAGCGAGTC
451 CAAGAATTG GACCCTCGGA TTCATAA

```

The PSORT algorithm predicts cytoplasm (0.4621).

The following *C.pneumoniae* protein (PID 4376558) was also expressed <SEQ ID 363; cp6558>:

```

1  MNIPAPQVPV IDEPVVNNTS SYGLSLKSSL RPITYLILAI LAIATLMSVL
51  YFCGIISVGT FVLGMLIPLS VCSVLCVAYL FYQQSSIEKT KVFSITSPSV
101 FFSDEDLNLL LGREEDSVSA IDELLKNFPA DDFRRPKMLP YSNFLDEQGR
151 PNESREEDSH TSKIL*

```

The cp6558 nucleotide sequence <SEQ ID 364> is:

```

1  ATGAACATAC CCGCTCCCCA AGTACCAGTC ATAGATGAGC CTGTAGTGAA
51  CAACACAAGT AGCTATGGTC TTTCATTGAA AAGTAGTTTA AGACCGATTA
101 CTTATTTGAT TTTAGCTATC TTAGCTATAG CCACACTGAT GTCTGTTCTC
151 TACTTTTGTG GCATCATTAG TGTGGGACG TTTGTTTGG GCATGCTGAT
201 CCTCTATCG GTCTGCTCTG TTCTTTGCGT TGCCTATTTA TTCTATCAGC
251 AATCTTCTAT AGAAAAGACT AAGGTCTTTT CTATAACCAG TCCTTCAGTA
301 TTTTCTCTG ATGAGGATCT TAATTTACTC TTAGGTCGAG AAGAAGATTC
351 AGTGTCTGCA ATTGATGAAC TTCTTAAGAA CTTTCCAGCT GATGATTTC
401 GTAGGCCGAA GATGCTTCCT TATTCAAATT TTCTAGATGA GCAGGGAAGG
451 CCTAATGAGA GTAGGGAAGA AGACTCTCAT ACTTCCAAGA TCTTATAA

```

The PSORT algorithm predicts inner membrane (0.4630).

The following *C.pneumoniae* protein (PID 4376630) was also expressed <SEQ ID 365; cp6630>:

```

1  MSMTIVPHAL FKNHCECHST FPLSSRTIVR IAIASLFCIG ALAALGCLAP
51  PVSIVGVSFV AFIAFVILSL VILALIFGEK KLPPTPRIIP DRFTHVIDEA
101 YGLSISAFVR EQQVTLAEFR QFSTALLCNI SPEEKIKQLP SELRSKVESF
151 GISRLAGDLE KMNWPIFEDL LSQTCPLYWL QKFISAGDPQ VCRDLGVPRE
201 CYGYWLGPL GYSTAKATIF CKETHHILQQ LTKEDVLLK NKALQEKWDT
251 DEVKAIVERI YTTYTARGTL KTEAGGLTKE TISKELLLS LHGYSFDQLQ
301 LITQLPRDAW DWLCFVDNST AYNLQLCALV GALSSQNLLD ESSIDFDVNL
351 GLYVIQDLKE AVQAFSASDE PKKELGKFL RHLSSVSKRL ESVLRQGLHR
401 IALEHGNARA RYVDVNFVTG ARIHRKTSIF FKD*

```

The cp6630 nucleotide sequence <SEQ ID 366> is:

```

1  ATGAGCATGA CGATCGTTCC ACATGCTTTA TTTAAAAATC ATTGCGAGTG
51  TCATTCTACC TTTCTTTTGA GTTCAAGGAC TATTGTAAGA ATAGCCATTG
101 CCGACCTCTT TTGTATAGGT GCATTAGCAG CTTTAGGCTG TTTGGCTCCT
151 CCGGTTTCTT ATATTGTTGG GAGTGTTTTA GCTTTTATTG CCTTTGTCAT
201 TCTTTCTTTA GTAATTTTAG CTTTGATTTT TGGAGAGAAG AAGCTTCCAC

```

-181-

1 MATSVAPSPV PESSPLSHAT EVLNLPNAYI TQPHPIPAAP WETFRSKLST  
 51 KHTLCFALTL LLTLGGTISA GYAGYTCNWI ICGIGLGIIIV LTLILALLLA  
 101 IPLKNKQTGT KLIDEISQDI SSIGSGFVQR YGLMPSTIKS VHLPELTTQN  
 151 QEKTRILNEI EAKKESIQNL ELKITECQNK LAQKQPKRKS SQKSFMRSIK  
 201 HLSKNPVILF DC\*

The cp6723 nucleotide sequence <SEQ ID 358> is:

1 ATGGCAACTT CCGTAGCCCC ATCACCAGTC CCCGAGAGCA GCCCTCTCTC  
 51 TCATGCTACA GAAGTTCTCA ATCTTCCTAA TGCTTATATT ACGCAGCCTC  
 101 ATCCGATTCC AGCGGCTCCT TGGGAGACCT TTCGCTCCAA ACTTTCCACA  
 151 AAGCATACGC TCTGTTTTCG CTTAACACTA CTGTTAACCT TAGGGGGAAC  
 201 GATCTCAGCA GGTACGCAG GATATACTGG AAACCTGGATC ATCTGTGGCA  
 251 TCGGCTTGGG AATTATCGTA CTCACACTGA TTCTTGCTCT TCTTCTAGCA  
 301 ATCCCTCTTA AAAATAAGCA GACAGGAACA AAACCTGATTG ATGAGATATC  
 351 TCAAGACATT TCCTCTATAG GATCAGGATT TGTTCTAGAGA TACGGGTTGA  
 401 TGTTCCTCTAC AATTAAAAGC GTGCATCTTC CAGAGCTGAC AACACAAAAT  
 451 CAAGAAAAAA CAAGAATTTT AAATGAAATT GAAGCGAAAA AGGAATCGAT  
 501 CCAAAATCTT GAGCTTAAAA TTAGTGAAGT CCAAAACAAG TTAGCACAGA  
 551 AACAGCCGAA ACGGAAATCA TCTCAGAAAT CATTTATGCG TAGTATTAAG  
 601 CACCTCTCCA AGAACCCTGT AATTTTGTTT GATTGCTGA

The PSORT algorithm predicts inner membrane (0.6095).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 179A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 179B) and for FACS analysis.

These experiments show that cp6723 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 180

The following *C.pneumoniae* protein (PID 4376749) was expressed <SEQ ID 359; cp6749>:

1 MSYYFSLWYL KVQHFQAAF DFTRSLCSRI SNFALGVIAL LPIIGQLYVG  
 51 LDWLLSRIKK PEFPSDVDQI VRVEHVVGHD HRSRVEDILK RQRLSLEPRD  
 101 EGKVHGDLPD APFF\*

The cp6749 nucleotide sequence <SEQ ID 360> is:

1 ATGAGTTATT ACTTTTCTCT TTGGTATCTG AAGGTGCAAC AGCACTTTCA  
 51 AGCAGCATTT GATTTTACTC GCTCCCTGTG TTCACGAATT TCTAATTTTG  
 101 CTTTGGGAGT GATTGCATTG CTTCCCTATTA TTGGGCAGTT GTATGTAGGG  
 151 CTGGACTGGC TCCTCTCTAG GATAAAAAAG CCAGAATTTT CTTCCGATGT  
 201 GGATCAGATC GTGCGAGTAG AACACGTCGT GGGTCACGAC CATAGAAGTC  
 251 GAGTTGAAGA TATTCTAAAG AGACAAAGGC TCTCATTAGA GCCTAGAGAC  
 301 GAGGGGAAGG TTCACGAGGA TCTGCCTTCA GCTCCTTTT TTTGA

The PSORT algorithm predicts inner membrane (0.2996).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 180A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 180B) and for FACS analysis.

These experiments show that cp6749 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

```

51  CACTAAAAAA AAATCCTGCA GCAACTTTGA TAAGATTCAG TCTCGAATTC
101 TATTGATTAC TGCAATCTTT GCTGTCTTAG TTACTATAGG GACCCTACTT
151 ATTGGTTTGC TTTTAAATAT TCCTGTTATC TATTTCTCTCA CAGGAATTTTC
201 ATTTATTGCT GTTGTCTTCA GCAACTTTAT CCTTTATAAA CGAGCAACCA
5  251 CCCTCTTAAA ACCGCGTGCT TGTGGCAAAC ACAAAGAAAT AAAACCAAAA
301 AGGGTCTCCA CCAACCTACA GTATTCTTCT ATCTCTATCG CAATCAATCG
351 TTCTAAAGAA AACTGGGAAC ACCAACCCTA GGACCTACAG AATCTCCCG
401 CACCCTCTGC ATTACTCACA GATAACCCTT ACGAGATATG GAAAGCTAAA
10 451 CATTCACTGT TTTCCCTAGT ATCCCTCCTA CCGGGAGGCA ATCCAGAACA
501 TCTCTTAATT TCAGCTTCCG AAAATTTTAG AAAGACTCTG TTAATTGAAG
551 AAACCTCGCA AAATGCGCCT ATATCCTCCT ACGTAGATAC CACTCCCCTC
601 CCAAAATCCT TGCTCAATGA GGCAATTCAG GAAACCAGGG TAGAAATAAA
651 TACAGAATC CCTGCGGAG ATTCAAGAGA ACGTTTATAC TGGCAACCCG
701 ATTTCCGAGG CCGCGTCTTC CTCCACAAA TACCAACAAC TCCTGAAGCC
15 751 ATCTACCAAT ACTACTATGC ACTCTATGTC ACTTATATCC AGACTGCGAT
801 CAATACGAAC ACCCAAATTA TCCAAATCCC TTTATACAGC TTGAGGGAGC
851 ATCTCTATTC TAGAGAATTG CCCCCGCAAT CAAGAATGCA ACAATCTTTG
901 GCTATGATTA CAGCAGTAAA ATACATGGCC GAGCTGCACC CAGAATATCC
20 951 GCTAACTATT GCTTGTGTTG AAAGATCCTT AGCCCAACTA CCTCAAGAAA
1001 GTATTGAGGA TCTCTCTTAG

```

The PSORT algorithm predicts inner membrane (0.5288).

The proteins were expressed in *E.coli* and purified as GST-fusion products. The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 181-185) and for FACS analysis.

25 These experiments show that cp6301, cp6558, cp6630, cp6633 and cp6642 are surface-exposed and immunoaccessible proteins, and that they are useful immunogens. These properties are not evident from their sequences alone.

### Example 186

The following *C.pneumoniae* protein (PID 4376389) was expressed <SEQ ID 371; cp6389>:

```

30 1  MSEVKPLFLK NDSFDLATQR FQNLINMLQE QAEIYNEYEE KNARVQNEIK
51  EQKDFVKRCI EDFEARGLV LKEELASLTR DFHDKAKAET SMLIECPCIG
101 FYYSIHQEEQ RQRQERLQKM AERYRDCQV LEAVQVEQKD MISSRVVDD
151 SYFEEKEEQ KVDNRKKEQD *

```

The cp6389 nucleotide sequence <SEQ ID 372> is:

```

35 1  ATGTCAGAAG TGAAGCCTTT GTTTTAAAG AATGACTCTT TTGATTTGGC
51  AACTCAGAGA TTCCAGAATC TAATTAACAT GCTACAAGAG CAAGCCGAGA
101 TATATAACGA GTATGAAGAA AAGAATGCTA GGGTTCAGAA TGAGATTAAG
151 GAGCAAAAGG ACTTTGTGAA AAGATGCATA GAGGACTTTG AAGCCAGAGG
201 ACTGGGGGTG CTAAGAAGAA AGCTTGCAAT TTTGACGCGT GATTTCCATG
40 251 ATAAAGCAAA AGCAGAGACT TCTATGCTCA TTGAATGTCC TTGTATTGGT
301 TTTTATTATA GTATTTCATCA GGAGGAACAA AGGCAAAGGC AAGAAAGGCT
351 TCAAAAAGATG GCTGAGCGCT ATAGGGACTG TAAACAAGTC TTGGAGGCTG
401 TCCAGGTGGA GCAAAAAGAT ATGATATCTT CTAGAGTCGT TGTCGATGAC
45 451 AGCTACTTTG AAGAAGAAAA AGAAGAACAA AAGGTGGATA ACAGAAAGAA
501 AGAACAGGAC TAG

```

The PSORT algorithm predicts cytoplasm (0.3193).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 186A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 186B) and for FACS analysis.

-183-

251 CAACACCAAG AATCATTCCCT GATAGATTTA CTCACGTGAT AGATGAAGCT  
 301 TATGGCCTTT CAATCTCTGC ATTTGTAAGA GAACAGCAGG TAACATTAGC  
 351 CGAGTTTAGA CAATTTTCTA CTGCCCTGTT GTGTAACATA TCTCCTGAAG  
 401 AGAAAATCAA ACAATTGCCT TCTGAATTGC GAAGTAAAGT AGAGAGTTTT  
 451 GGTATTAGCA GGCTCGCAGG TGATTTAGAA AAGAATAATT GGCCAATATT  
 501 TGAAGATCTT TTAAGCCAAA CCTGCCCGTT ATATTGGCTT CAGAAATTTA  
 551 TATCAGCAGG AGATCCACAA GTTTGTAGAG ACCTAGGTGT CCCTAGAGAA  
 601 TGTTATGGGT ACTATTGGCT AGGGCCTTTG GGATACAGTA CAGCTAAGGC  
 651 TACAATTTTT TGTAAGAGA CGCATCATAT TCTTCAACAA TTAACGAAAG  
 701 AGGACGTTCT TTTATTAAAA AACAAGGCTC TTCAAGAGAA ATGGGATACT  
 751 GATGAAGTCA AAGCAATTGT AGAGCGTATC TACACTACCT ATACGGCACG  
 801 AGGAACCTCA AAGACCGAAG CAGGGGGACT TACAAAAGAG ACAATCAGTA  
 851 AGGAATTGCT ATTGTTGAGC TTGCATGGCT ATTCTTTTGA TCAGTACAG  
 901 CTGATCACTC AACTTCCTAG AGATGCTTGG GATTGGCTGT GTTTTGTAGA  
 951 TAACAGTACC GCATACAACC TTCAGCTTTG TGCTCTTGTA GGAGCTTTGT  
 1001 CATCCCAAAA TCTTCTTGAC GAATCTTCTA TCGATTTTGA TGTAAACCTA  
 1051 GGCCTGTATG TGATTCAAGG TCTAAAAGAA GCTGTTCAAG CATTTTCTGC  
 1101 TTCTGATGAG CCAAAGAAAG AACTAGGTAA ATTCTTGTTA AGGCATTTGA  
 1151 GTTCAGTTTC TAAGCGATTA GAGAGTGAT TAAGACAGGG TCTTCACAGA  
 1201 ATAGCTCTAG AGCATGGAAA TGCCAGAGCT AGGGTTTATG ACGTCAATTT  
 1251 TGTAACAGGA GCTAGAATTC ATAGGAAGAC GAGTATCTTC TTAAAGACT  
 1301 AA

The PSORT algorithm predicts inner membrane (0.7092).

The following *C.pneumoniae* protein (PID 4376633) was also expressed <SEQ ID 367; cp6633>:

25 1 MVNIQPVYRN TQVNYSQATQ FSVCPALSL IIVSVVA AVL AIVALVCSQS  
 51 LLSIELGTAL VLVSLILPAS AMFMIYKMRQ BPKELLIPKK IMELIQEHYP  
 101 SIVVDFIRDQ EVSIYEIHL ISILNKTNVF DKAPVYLQEK LLQFGIEKFK  
 151 DVHPSKLPNF EEILLQHCPL HWLGRVLVPM VSDVTPGTYG YYWCGPLGLY  
 201 ENAPSLFERR SLLLLKKISF GEFALLEDLG KKNWSSSEL VQIRQNLFR  
 30 251 YYADKEEVDE AELNADYEQF DSSLHLIFSH KLS\*

The cp6633 nucleotide sequence <SEQ ID 368> is:

1 ATGGTTAATA TACAGCCTGT GTATAGGAAT ACCCAAGTCA ACTATAGTCA  
 51 GGCTACCCAA TTTTCGGTGT GCCAGCCAGC GCTTAGCCTG ATTATCGTTT  
 101 CTGTTGTTGC TGCTGTACTC GCTATTGTAG CTTTGGTATG CAGTCAATCT  
 35 151 CTTTTATCCA TAGAGTTAGG AACTGCTCTT GTTCTAGTTT CTCTTATTCT  
 201 TTTTGCTTCT GCTATGTTTA TGATTATATA GATGAGACAA GAACCTAAGG  
 251 AGTTGCTGAT CCCTAAGAAA ATCATGGAAC TCATCCAAGA ACATTATCCA  
 301 AGTATTGTTG TTGATTTTAT TAGAGATCAG GAGGTTTCCA TTTATGAGAT  
 351 ACATCACTTG ATCTCTATTC TTAATAAGAC GAATGTTTTT GACAAAGCAC  
 40 401 CAGTATATTT ACAAGAAAAA CTCTTACAGT TTGGCATTGA GAAGTTCAAA  
 451 GATGTACATC CAAGTAAGCT CCCTAATTTT GAAGAAATTC TTCTACAGCA  
 501 TTGCCCATG CATTGGTTGG GACGCTGGT ATATCCCATG GTATCGGATG  
 551 TCATCCAGG AACCTATGGA TACTATTGGT GTGGTCCTTT AGGACTGTAC  
 601 GAGAACGCTC CCTCTCTTTT TGAACGTCGA TCTCTTCTAT TGTAAAGAA  
 45 651 AATTAGCTTT GGAGAGTTTG CTCTTTTAGA AGATGGTCTC AAGAAAAACA  
 701 CGTGGAGTTC TTCGGAATC GTTCAAATCA GACAAAACCT TTTTACAAGA  
 751 TATTATGCTG ATAAAGAAGA GGTAGATGAA GCAGAGTTAA ACGCTGATTA  
 801 CGAACAGTTT GATCCCTCC TTCACCTTAT TTTTCTCAC AAGCTCTCTT  
 851 GA

50 The PSORT algorithm predicts inner membrane (0.7283).

The following *C.pneumoniae* protein (PID 4376642) was also expressed <SEQ ID 369; cp6642>:

1 MATISPISLT VDHPLVDTKK KSCSNFDKIQ SRILLITAIF AVLVTIGTLL  
 51 IGLLLNIPVI YFLTGISFIA VVLSNFILYK RATTLLKPRA CGKHKEIKPK  
 101 RVSTNLQYSS ISIAINRSKE NWEHQPKDLQ NLPAPSALLT DNPYEIWKAK  
 55 151 HSLFSLVSL PGGNPEHLI SASENLGKTL LIEETSQNAF ISSYVDTPPS  
 201 PKSLLEAIQ ETRVEINTEL PAGDSGERLY WQPDFRGRVF LPQIPTTPEA  
 251 IYQYYYALYV TYIQTAINTN TQIIQIPLYS LREHLYSREL PPQSRMQQSL  
 301 AMITAVKYMA ELHPEYPLTI ACVERSLAQL PQESIEDLS\*

The cp6642 nucleotide sequence <SEQ ID 370> is:

60 1 ATGGCTACAA TCTCACCCAT ATCTTTAACT GTAGATCATC CCCTAGTAGA

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 188A; lanes 2-3). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 188B) and for FACS analysis.

These experiments show that cp6868 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 189

The following *C.pneumoniae* protein (PID 4376894) was expressed <SEQ ID 377; cp6894>:

```

1  MYKRCVLDKI LKGIVAGSLI LLYWSSDLLE RDIKSIKGNV RDIQEDIREI
51  SRVVKQQQTS QAIPAAPGVM LAPKLVRDEA FALLFGDPSY PNLLSLDPYK
101 QQTLPELLGT NFHPHGILRT AHVGKPENLS PFNGFDYVVG FYDLCIPSLA
151 SPHVKGKYEEL SPDLAVKIEE HLVEDGSGDK EFHIYLRPNV FWRPIDPKAL
201 PKHVQLDEVF QRPHPVTAHD IKFFYDAVMN PYVATMRAVA LRSCYEDVVS
251 VSVENDLKLK VRWKAHTVIN EEGKEERKVL YSAFSNTLSL QPLPRFYQYQ
301 FANGEKIIED ENIDTYRTNS IWAQNFTMHW ANNYIVSCGA YYFAGMDDEK
351 IVFSRNPDFY DPLAALIDKR FVYFKESTDS LFQDFKTGKI DISYLPNQR
401 DNFYSFMKSS AYNKQVAKGG AVRETVSADR AYTYIGWNCF SLFFQSRQVR
451 CAMNMAIDRE RIIEQCILDGQ GYTISGPFAS SSPSYNKQIE GWHYSPEEAA
501 RLLEEEGWID TDGDGIREKV IDGVIVPFRF RLCYYVKSVT AHTIADYVAT
551 ACKEIGIECS LLGLDMADLS QAFDEKNFDA LLMGWCLGIP PEDPRALWHS
601 EGAMEKGSAN VVGFFHNEAD KIIDRLSYEY DLKERNRLYH RFHEIIHEEA
651 PYAFLFSRHC SLLYKDYVKN IFVPTHRTDL IPEAQDETVN VTMVWLEKKE
701 DPCLSTS*

```

The cp6894 nucleotide sequence <SEQ ID 378> is:

```

1  ATGTATAAAA GATGTGTGCT AGATAAAATT TTAAAGGGGA TTGTCGCCGG
25 51  TTCTTTAATT TTGTTATACT GGTCCCTCAGA CCTACTTGAA AGAGACATTA
101 AGTCGATAAA AGGTAACGTA AGAGATATTC AAGAAGACAT TCGTGAAATC
151 TCACGCGTAG TGAAACAACA GCAGACATCA CAAGCTATCC CTGCGGCACC
201 TGGGGGTGATG CTCGCTCCTA AGCTCGTCAG AGACGAAGCT TTTGCTCTAC
251 TCTTTGGAGA TCCTAGTTAT CCTAATTTAC TTTCCCTAGA CCCCTATAAA
30 301 CAGCAGACTC TTCCTGAACT TCTAGGAACA AATTTCCACC CTCATGGTAT
351 CCTACGCACT GCCCATGTCG GAAAACCCGA AAATCTGAGC CCTTTTAAATG
401 GCTTTGATTA TGTCGTGGGC TTTTACGATC TCTGTATTCC TAGTTTAGCT
451 TCTCCCCACG TAGGGAATA CGAAGAATT TCTCCAGATC TCGCTGTGAA
501 AATAGAAGAA CATCTTGTG AAGATGGTTC TGGGGATAAA GAGTTTCACA
35 551 TCTATCTGAG GCCGAATGTT TTTTGGCGTC CTATAGATCC TAAGGCCCTT
601 CCAAAACACG TTCAGTTAGA CGAAGTATTT CAACGTCCTC ATCCTGTGAC
651 AGCTCATGAT ATTAAGTTT TCTACGACGC TGTTATGAAC CCTTATGTAG
701 CAACCATGCG AGCAGTGGCT CTGCGCTCTT GTTATGAAGA TGTGGTTTCT
751 GTCTCAGTAG AAAACGATTT AAAATTAGTA GTCAGATGGA AAGCACACAC
40 801 GGTAAATCAAT GAAGAAGGAA AGGAAGAGCG CAAAGTGCTC TACTCTGCAT
851 TTTCTAATAC CTTAAGCTTG CAGCCCCCTC CTAGATTGTG ATATCAGTAT
901 TTTGCTAACG GGGAAAAAAT CATTGAAGAT GAGAATATCG ATACCTACCG
951 AACCAATTCC ATTTGGGCGC AAAACTTCAC TATGCATTGG GCAAACAAC
45 1001 ATATTGTAAG TTGTGGAGCC TACTACTTTG CAGGGATGGA TGATAGAGAA
1051 ATCGTGTTTT CTAGAAATCC TGACTTCTAT GATCCTCTTG CGGCTCTTAT
1101 TGACAAGCGT TTCGTCTATT TTAAGGAAAG CACAGACTCC CTATTCCAAG
1151 ATTTTAAGAC AGGGAAAAA GACATCTCTT ACCTTCCACC CAACCAAGA
1201 GATAATTTCT ATAGTTTTAT GAAAAGCTCC GCTTATAACA AACAGGTAGC
1251 TAAGGGAGGA GCCGTCCGTG AAACAGTCTC AGCAGATCGA GCATATACGT
50 1301 ACATAGGATG GAATTGCTTT TCATTATTTT TCCAAAGCCG ACAGGTGCGC
1351 TGTGCTATGA ACATGGCAAT CGATAGAGAG AGGATTATCG AACAGTCTT
1401 GGATGGCCAA GGCTATACGA TTAGTGGGCC TTTTGCTTCG AGTTCCTCCT
1451 CTTATAATAA ACAGATCGAA GGGTGCAATT ATTCTCCAGA AGAAGCAGCT
1501 CGTCTCCTGG AAGAAGAGGG ATGGATAGAT ACCGATGGCG ATGGAATCCG
55 1551 AGAAAAAGTT ATCGATGGTG TGATTGTCCC GTTCCGTTTC CGTTTATGCT
1601 ATTATGTAAG GAGTGTACAC GCTCATACCA TTGCAGATTA CGTAGCTACT
1651 GCTTGTAAGG AAATCGGAAT CGAGTGTAGC CTTCTAGGAC TAGATATGGC
1701 CGATCTTTTC CAAGCTTTTG ATGAAAAGAA TTTTCGATGCT CTTTTAATGG
1751 GATGGTGTTC AGGAATTCCT CCTGAGGATC CTAGGGCTTT ATGGCATTCT

```



These experiments show that cp6389 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 187

The following *C.pneumoniae* protein (PID 4376792) was expressed <SEQ ID 373; cp6792>:

```

5      1  VLQEHFFLSE DVITLAQQLL GHKLITHEG LITSGYIVET EAYRGPDDKA
      51  CHAYNYRKTQ RNRMYLKGG SAYLYRCYGM HHLNVVTGP EDIPHAVLIR
     101  AILPDQKEL MIQRRWRDK PPHLLTNGPG KVCQALGISL ENNRQRLNTP
     151  ALYISKEKIS GTLTATARIG IDYAQYRDV PWRFLSPED SGKVL*

```

The cp6792 nucleotide sequence <SEQ ID 374> is:

```

10      1  GTGCTACAAG AACATTTTTC TCTATCGGAA GATGTAATTA CACTAGCGCA
      51  ACAGCTTTTA GGACATAAAC TCATCACAAC ACATGAGGGT CTGATAACTT
     101  CAGGTTACAT TGTAGAAACC GAAGCGTATC GTGGCCCTGA TGACAAAGCA
     151  TGCCACGCCT ACAACTACAG AAAAAGTCAG AGGAACAGAG CGATGTACCT
     201  GAAAGGAGGC TCTGCTTACC TCTACCGTTG CTATGGCATG CATCACCTAT
15      251  TGAATGTTGT CACTGGACCT GAGGACATTC CCCATGCCGT CCTGATCCGG
      301  GCCATCCTTC CTGATCAAGG CAAAGAACTT ATGATCCAAC GCCGCCAATG
     351  GAGAGATAAA CCCCCACACC TTCTCACCAA TGGACCCGGA AAAGTGTGCC
     401  AAGCTCTAGG AATCTCTTTG GAAAACAATA GGCAACGCCT AAATACCCCA
     451  GCTCTCTATA TCAGCAAAGA AAAAATCTCT GGGACTCTAA CAGCAACTGC
20      501  CCGGATCGGC ATCGATTATG CTCAAGAGTA TCGTGATGTC CCATGGAGAT
     551  TTCTCCTATC CCCAGAAGAT TCGGGAAAAG TTTTATCTTA A

```

The PSORT algorithm predicts cytoplasm (0.180).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 187A; lanes 2-4).

The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 187B) and for FACS analysis.

These experiments show that cp6792 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 188

The following *C.pneumoniae* protein (PID 4376868) was expressed <SEQ ID 375; cp6868>:

```

30      1  MVETVLHNFQ RYLSKYLYRV FRFPCRKKTFF LSSHRVLARP SFPVDYCPGK
      51  IYDLQEIYEE LNAQLFQGAL RLQIGWFGRK ATRKGKSVVL GLFHENEQLI
     101  RIHRSLDRQE IPRFFMEYLV YHEMVHSVVP REYSLSGRSI FHGKKFKEYE
     151  QRFPLYDRAV AWEKANAYLL RGYKKRVGGG YGRA*

```

The cp6868 nucleotide sequence <SEQ ID 376> is:

```

35      1  ATGGTTGAAA CAGTACTTCA TAATTTCCAA CGTTATCTGA GCAAGTATCT
      51  CTATAGGGTA TTTTCGTTCC CATGTCGTAA AAAGACGTTT CTATCTTCGC
     101  ACAGGGTTCT TGCTCGTCCT TCATTCCCAG TAGACTACTG TCCGGGAAAG
     151  ATCTATGATT TGCAGGAGAT CTATGAGGAA TTGAATGCGC AGTTATTTCA
     201  AGGTGCACTG CGTTTACAGA TTGGTTGGTT CGGAAGGAAA GCTACCAGAA
40      251  AAGGCAAGAG TGTGTCTTTG GGATTGTTTC ATGAAAATGA ACAGTTAATT
      301  CGAATTCATC GTTCTTTAGA TCGGCAGGAA ATCCAAGAT TTTTATGGA
     351  ATATCTTGTG TATCATGAAA TGGTTCATAG TGAGTCCCT AGAGAGTATT
     401  CTCTATCGGG GCGTTCGATT TTTTATGGTA AAAAGTTTAA AGAATACGAA
     451  CAACGTTTCC CCTTGTATGA TCGTGTGTT GCTTGGGAAA AGGCAAACGC
45      501  TTATTTATTG CGAGGTATA AAAAAAGAGT AGGTGGAGGA TATGGCAGGG
     551  CATAG

```

The PSORT algorithm predicts bacterial cytoplasm (0.325).

TABLE II – sequences of the primers used to amplify Cpn genes.

Orf ID	N-terminus final primer	C-terminus final primer
CP0014P	GCGTC CCG GGT CATATG AAGTCTTCTTTCCCA	GCGT CTC GAG ATGAAAGAGTTTTCGG
CP0015P	GCGTCCCGGGTCATATG TCAGTCTCTTTTCTGA	GCGT CTC GAG GAATTGGTATTTTGCTC
CP0016P	GCGTCCCGGGTCATATG GCCGATCTCACATTAG	GCGT CTC GAG GTCCAAAGTTAAGGTAGCA
CP0017P	GCGT CCG GGT CATATG GGTATCAAGGGAAGT	GCGT CTC GAG AAATCCGAATCTTCC
CP0019P	GCGTCCCGGGTCAT ATGCAAGACTCTCAAGACTATAG	GCGT CTC GAG AAATCGGTATTTACCC
CP6260P	GCGTC CCG GGT GCTAGCACTACGATTTCTTTAACCC	GCGT CTC GAG AAAACGAAATTTGCTTC
CP6397P	GCGTC CCG GGT CATATGTTTAACTGCTAAAAATCTATT	GCGT CTC GAG ATGAAAGAAGAGTCCCTCG
CP6456P	GCGTC CCG GGT CATATG TCATCTCCTGTAATAACA	GCGT CTC GAG CTGACCATCTCTCTGT
CP6466P	GCGTC CCG GGT CAT ATG TGCAAGGAGTCCAGT	GCGT CTC GAG ATTTTCCTTAGCATAACG
CP6467P	GCGTC CCG GGT CAT ATG TGTTCCCATCCCAA	GCGT CTC GAG TAGTTTTTCTATAAAACGAAAGTCT
CP6468P	GCGTC CCG GGT CAT ATG TGCTCCTCTACTCTTC	GCGT CTC GAG GGGGAAATAGGTATATTGA
CP6469P	GCGTC CCG GGT CAT ATG AGCTGCTCAAAGCAA	GCGT CTC GAG ACTTAAGATATCGATATTTTGA
CP6552P	GCGTC CCG GGT CAT ATG TGCCATAAGGAAGATG	GCGT CTC GAG ACCATTGTCTTGAGTCAT
CP6567P	GCGTC CCG GGT CAT ATG ACCTACCGATCCCC	GCGT CTC GAG AGAAGCCGGTAGAGGC
CP6576P	GCGTC CCG GGT CAT ATG ACTGAAAAAGTTAAAGAAGG	GCGT CTC GAG GAA CATGCCCCCTAA
CP6727P	GCGTC CCG GGT CATATGCTACATCCACTAATGGC	GCGT CTC GAG GAAAGAATAACGAGTTCC
CP6729P	GCGTC CCG GGT CAT ATGCGAGATGCTTCTTTATC	GCGT CTC GAG GAATGAGTATCTTAGCC
CP6731P	GCGTC CCG GGT CATATGCTGTGTGTAATAAATCAAT	GCGTC CAT GGC GGC CGC GAACTGGAACTTACCTCC
CP6736P	GCGTC CCG GGT GCT AGCGTAGAAGTTATCATGCCTT	GCGTC CAT GGC GGC CGC AAATCGTAATTTGCTTC
CP6737P	GCGT GGA TCC CAT ATG GAGACTAGACTCGGAGG	GCGT CTC GAG AAATGTGGATTTTAGTCC
CP6751P	GCGTC CCG GGT GCT AGC AATGAAGGTCTCCAACT	GCGT CTC GAG AAATCTCATCTACTCGC
CP6752P	GCGTGA ATT CAT ATGTTCCGGATGACTCCT	GCGT CTC GAG GAATTTTAAGGTACTTCTCG
CP6753P	GCGTC CCG GGT GCT AGCACTCCCTACTCTCATAGAG	GCGT CTC GAG AAACCTAAAGGTCTGTT
CP6787P	GCGTC CCG GGT CAT ATG ATAAACAATAAGGCCGT	GCGT CTC GAG TTCGTAAGCAACTTCAGA
CP6829P	GCGTC CCG GGT CAT ATG AAGCAGATGCGTCTT	GCGTC CAT GGC GGC CGC GAAACTAAGGAGAGGC
CP6830P	GCGTC CCG GGT CAT ATG GATCCCGCGTCTGTT	GCGTC CAT GGC GGC CGC GAATACAAACCGGATCC
CP6832P	GCGTC CCG GGT CAT ATG CATAAAGTAATAGTTTTCATT	GCGT CTC GAG TAACTAGAAAAAGTCGTC
CP6848P	GCGTC CCG GGT CAT ATG TCATCAAACTACATCCC	GCGT CTC GAG AACGCGAGCTATTTTAC
CP6849P	GCGTC CCG GGT GCT AGC AGCGGGGTATAGAG	GCGT CTC GAG ATACACGTGGGTATTTTC
CP6850P	GCGTC CCG GGT CAT ATG TGCCGATGTAGAT	GCGT CTC GAG CTGTTTGATCTGACC
CP6854P	GCGTC CCG GGT GCT AGC TCAATAGCTATTGCAAG	GCGT CTC GAG TTATCGAAATGTCTTTG
CP6879P	GCGTC CCG GGT CAT ATG GCAACACCGCTCAA	GCGTC CAT GGC GGC CGC TCCTTGAATGTCTCTGC
CP6894P	GCGTC CCG GGT CAT ATG TATAAAGATGTGTGCTAGA	GCGT CTC GAG GGATGTACTTAAGCAG
CP6900P	GCGTC CCG GGT CAT ATG AAGATAAAATTTCTTGAAG	GCGT AAG CTT GGGAAAGACGATACCG
CP6952P	GCGTC CCG GGT CAT ATG CTCTCGGATCAATATATAGG	GCGT CTC GAG TCGAATTTCTTTTTTAGC
CP7034P	GCGTC CCG GGT CAT ATG AAAAAACAGGTATATCAATG	GCGT AAG CTT AAACGCTGAAATTTATACC
CP7090P	GCGTC CCG GGT CAT ATG TGTAGCCTTCCOCT	GCGT CTC GAG GCGTGCATGAATCTTA
CP7091P	GCGTC CCG GGT CAT ATG GAAGAATTAGAAGTTGTGT	GCGT CTC GAG TAGTGTCTCTTTATFCGGT
CP7170P	GCGTC CCG GGT CAT ATG CTAGGGGTGGAACC	GCGT AAG CTT AAACGCGAGACCTGACG
CP7228P	GCGTC CCG GGT CAT ATG ACTGCTGTCTTATTCTTACA	GCGT CTC GAG ATCTGAAAGCGGAGG
CP7249P	GCGTC CCG GGT CAT ATG ATCCCATCCCTACC	GCGT CTC GAG ATCAGGTTGCTGAGACTT
CP7250P	GCGTC CCG GGT CAT ATG AATCTTTCAAACAGGTCT	GCGT CTC GAG ATTTTCTTAGAGAGACTCTC
CP0018P	GTGCGT CATATG GCAACCACTCCACTAA	ACTCGCTA GCGGCCGC TAATGAGGTCCCCAG
CP6270P	GTGCGT CATATG AATTTATTAGGAGCTGCT	ACTCGCTA GCGGCCGC AAATTTGATTTTGCTACC
CP6735P	GTGCGT CATATG GCAGCACAGTTGTATAT	ACTCGCTA GCGGCCGC TGGCGTAGAAGTGATC
CP6998P	GTGCGT CATATG TTGCTGTAGGGAAC	ACTCGCTA GCGGCCGC GAATCTGAAGTACCCAGA
CP7033P	GTGCGT CATATG GTTAATCTTATTTGTTCCA	ACTCGCTA GCGGCCGC TTGGAGATAACAGAAATATA
CP7287P	GTGCGT CATATG TTACACAGCTCAGAACTAGA	ACTCGCTA GCGGCCGC GAAAAATAACGGATACCA
CP0010P	GTGCGT CATATG GCAACTGCTGAAAAATATA	GCGT CTCGAG GAATTGGAATTTACCC
CP0468P	GTGCGT GCTAGC ATTTTATTATGACAACTCTAT	GCGT CTCGAG AAATGTGCAATGACTCT
CP6272P	GTGCGT CATATG TTGACTCATCAAGAGGCT	GCGT CTCGAG GAAGGGAGGTTTTTTAGGT
CP6273P	GTGCGT CATATG ACATATCTGGAAGCTC	ACTCGCTA GCGGCCGC CTCCACAATTTTATG
CP6362P	GTGCGT CATATG CCTTTGATATTACTTATTATACA	GCGT CTCGAG TCGTTTCCAAATCCA
CP6372P	GTGCGT CATATG AAACAACACTATTCTCTAAATA	GCGT CTCGAG TTTCTTGTGGTTTTTCT
CP6390P	GTGCGT CATATG CGAGAGGTGCCTAAG	ACTCGCTA GCGGCCGC TCTCTAGACAGCCTT
CP6402P	GTGCGT CATATG AATGTTCCGGAATCTCCTTT	GCGT CTCGAG GAAGGGGTTGGCCGT
CP6446P	GTGCGT CATATG TGTAAATCAAAAGCCCTCTT	GCGT CTCGAG GGGCTGAGGAGGAAC
CP6520P	GTGCGT GCTAGC AAACACTACCTATCATTTTCT	GCGT CTCGAG CAGAAAGGCTTTTCTTT
CP6577P	GTGCGT CATATG AATTTAGGCTATGTTAATTTA	GCGT CTCGAG GTTTTGTTTTGTGAAGA
CP6602P	GTGCGT CATATG GCAGCATCAGGAGGCA	GCGT CTCGAG TGACCAAGGATAGGGTTAG

-187-

1801 GAAGGGGCTA TGGAAAAGGG TTCAGCGAAT GTTGTAGGTT TCCATAATGA  
 1851 AGAAGCTGAT AAAATCATAG ACAGACTCAG CTACGAATAC GATCTGAAAG  
 1901 AACGTAATCG CCTGTACCAC CGTTTCCATG AAATTATTCA TGAGGAAGCT  
 1951 CCTTATGCTT TCTTGTCTC ACGACATTGT TCCTTACTTT ATAAGGATTA  
 2001 TGTAAAAAAT ATTTTCGTAC CTACACATAG AACAGATTTA ATTCCTGAAG  
 2051 CTCAGGATGA GACTGTCAAC GTAACATATG TATGGCTTGA GAAGAAGGAG  
 2101 GATCCGTGCT TAAGTACATC CTAA

The PSORT algorithm predicts inner membrane (0.162).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 189A) and also in  
 10 GST/his form. The recombinant proteins were used to immunise mice, whose sera were used in a  
 Western blot (Figure 189B) and for FACS analysis.

These experiments show that cp6894 is a surface-exposed and immunoaccessible protein, and that it  
 is a useful immunogen. These properties are not evident from the sequence alone.

### Example 190

15 The following *C.pneumoniae* protein (PID 4377193) was identified in the 2D-PAGE experiment  
 <SEQ ID 379; cp7193>:

1 MKRVIYKTIF CGLTLLTSLS SCSLDPKGYN LETKNSRDLN QESVILKENR  
 51 ETPSLVKRLS RRSRRLFARR DQTQKDTLQV QANFKTYAEK ISEQDERDLS  
 101 FVVSSAAEKS SISLALSQGE IKDALYRIRE VHPLALIEAL AENPALIEGM  
 20 151 KKMQRDWIW NLFLTQLSEV FSQAWSQGI SEEDIAAFAS TLGLDSGTVA  
 201 SIVQGERWPE LVDIVIT\*

A predicted leader peptide is underlined.

The cp7193 nucleotide sequence <SEQ ID 380> is:

25 1 ATGAAAAGAG TCATTTATAA AACCATATTT TCGGGGTTAA CTTTACTTAC  
 51 AAGTTTGAGT AGTTGTTCCC TGGATCCTAA AGGATATAAC CTAGAGACAA  
 101 AAAACTCGAG GGACTTAAAT CAAGAGTCTG TTATACTGAA GGAAAACCGT  
 151 GAAACACCTT CTCCTGTAA GAGACTCTCT CGTCGTTCTC GAAGACTCTT  
 201 CGCTCGACGT GATCAAATC AGAAGGATAC GCTGCAAGTG CAAGCTAACT  
 251 TTAAGACCTA CGCAGAAAAG ATTTTCAGAGC AGGACGAAAG AGACCTTTCT  
 30 301 TTCGTTGTCT CGTCTGCTGC AGAAAAGTCT TCAATTTCTG TAGCTTTGTC  
 351 TCAGGGTGAA ATTAAGGATG CTTTGTACCG TATCCGAGAA GTCCACCCTC  
 401 TAGCTTTAAT AGAAGCTCTT GCTGAAAACC CTGCCTTGAT AGAAGGGATG  
 451 AAAAAGATGC AAGGCCGTGA TTGGATTGG AATCTTTTCT TAACACAATT  
 501 AAGTGAAGTA TTTTCTCAAG CTGGTCTCA AGGGGTTATC TCTGAAGAAG  
 35 551 ATATCGCCGC ATTTGCCTCC ACCTTAGGTT TGGACTCCGG GACCGTTGCG  
 601 TCCATTTGTC AAGGGGAAAG GTGGCCCGAG CTTGTGGATA TAGTGATAAC  
 651 TTAA

The PSORT algorithm predicts periplasmic (0.925).

This shows that cp7193 is an immunoaccessible protein in the EB and that it is a useful immunogen.

40 These properties are not evident from the protein's sequence alone.

It will be appreciated that the invention has been described by way of example only and that  
 modifications may be made whilst remaining within the spirit and scope of the invention.

CP7342P	GTGCGT CATATG AAAAAAAAAATTATTTCTACT	ACTCGCTA GCGGCCGC CACACTCTGTTCTTCTG
CP7347P	GTGCGT CATATG TTTTCTAAGGATTGACTAA	GCGT CTCGAG CGAAGCAGAAGTCGT
CP7353P	GTGCGT CATATG AATATGCCTGTTCTCTCT	GCGT CTCGAG GGGGCGTAGGTTGTA
CP7193P	GTGCGT CATATG TGTTCCTTGATCCT	ACTCGCTA GCGGCCGC AGTTATCACTATATCCACAAG
CP7248P	GTGCGT GCTAGC CTTGAACATTCATAACAAGAT	GCGT CTCGAG ACGTAGTTTAAGAGCAGACT
CP7261P	GTGCGT CATATG TGTCTATCTGCCTACATAG	GCGT CTCGAG TTTTGATGCTTCTTTCA
CP7280P	GTGCGT CATATG GACCAGAAAATTGAAAA	GCGT CTCGAG AGAGGTCTTCTGAGTGC
CP7302P	GTGCGT CATATG AATTTCATTTGTAGTGTAGT	GCGT CTCGAG GAACAGTTCGATTTGTG
CP7306P	GTGCGT CATATG CTTCCTTTATCAGGGCA	ACTCGCTA GCGGCCGC TTCTTCAGTTTCAGG
CP7367P	GTGCGT GCTAGC CGTTATGCCGAGGTC	GCGT CTCGAG TTCGTGCAATTTGGTG
CP7408P	GTGCGT CATATG TTGAAAATCCAGAAAAA	GCGT CTCGAG ATTCATTTTCGGAAGAG
CP7409P	GTGCGT CATATG AGACGTTATCTTTTCATGGT	GCGT CTCGAG CCTTTGCTCTTTACATAG
CP6733P	GTGCGT ACTAGT TGTCACCTACAGTCACTAG	GCGT CTCGAG GAATCGGACTTTGGTA
CP6728P	GTGCGT ACTAGT AAGTCCTCTGCTCTTGG	GCGT CTCGAG GAAACAAAACCTAGAGCCC

TABLE III – Proteins with best results in FACS analysis

cp number	Molecular Weight (kDa)		Fusion type
	Theoretical	Western Blot	
6260	97.5	94; 70	GST
6270	87.5	-	GST
6272	78.0	90	GST
6273	58.6	74; 64; 50	GST
6296	31.1	-	GST
6390	88.9	102	GST
6456	42.5	89; 67,45	GST
6466	57.5	59; 56	His
6467	59.0	67	GST
6552	28.4	50; 27	GST
6576	86.0	79; 70; 62; 45	GST
6577	17.3	12	GST
6602	43.4	53; 42; 34	GST
6664	54.5	104; 45	GST
6696	47.9	95; 53	GST
6727	130.0-142.9	123; 61; 39	His
6729	94.8	multiple bands	GST
6731	95.5	97	GST
6733	97.1	104	His
6736	100.1	98; 93; 66; 60	GST
6737	101.2	multiple bands	GST
6751	100.2	95; 71	GST
6752	102.1	97; 48	His
6767	29.1	28	GST
6784	32.9	35	GST
6790	71.3	multiple bands	His
6802	29.7	-	GST
6814	29.6	28	GST

CP6607P	GTGCGT	CATATG	CCTCGTGGTGACACTTT	GCGT	CTCGAG	CGCTGCTTCTTGCTC
CP6615P	GTGCGT	CATATG	TGCTCTCAAAAAACGACAA	GCGT	CTCGAG	TGAAGAGGCGCCATC
CP6624P	GTGCGT	CATATG	GATGCGAAAAATGGGA	GCGT	CTCGAG	TCTTTGACATTCAAGAGC
CP6672P	GTGCGT	CATATG	ATTCCCTACCATGTTAATG	GCGT	CTCGAG	GTCATACAAATTCCTTATATA
CP6679P	GTGCGT	CATATG	TGCACTCACTTAGGCT	GCGT	CTCGAG	CGAGTAGTTAGCACAAAC
CP6717P	GTGCGT	GCTAGC	AAGACAATCGTAGCTTCA	ACTCGCTA	GCGGCCGC	GGCTGGCATATAGGT
CP6784P	GTGCGT	GCTAGC	AAATCAAGATGTTCTATTGATA	GCGT	CTCGAG	TCCAAAACAACCCCTCT
CP6802P	GTGCGT	CATATG	TGCGTAAGTTATATTAATTCCTT	GCGT	CTCGAG	CAGTCGGGCTTGTTG
CP6847P	GTGCGT	CATATG	TCCGATCTTTTACGAG	GCGT	CTCGAG	TTTCTACACTGTTGTAATAAA
CP6884P	GTGCGT	CATATG	AATCAGCTGCTTTCT	GCGT	CTCGAG	AGAGAAGGTAATGTACC
CP6886P	GTGCGT	CATATG	TGTCTACTTATTATCTATCTCTAC	GCGT	CTCGAG	TTCAGAAAAATGGCT
CP6890P	GTGCGT	CATATG	TCCCCACGACGACAA	GCGT	CTCGAG	TCCTGCAGCATTAGC
CP6960P	GTGCGT	CATATG	TGTGACGTACGGTCTA	ACTCGCTA	GCGGCCGC	TTACCTTGATTTTCCT
CP6968P	GTGCGT	CATATG	TGCGATGCAAAAC	ACTCGCTA	GCGGCCGC	GGAAGTATGCTTAGATATT
CP6969P	GTGCGT	CATATG	TGCTGTGGTTACTCTATT	ACTCGCTA	GCGGCCGC	AAAAGGTCATAGTATACCT
CP7005P	GTGCGT	CATATG	AAAAGTGTGATATTGAACA	GCGT	CTCGAG	CTGAGCTTCTATTTCATTAT
CP7072P	GTGCGT	CATATG	CCCATTTATGGGAAA	GCGT	CTCGAG	GTGAGCAAGAGTTTG
CP7101P	GTGCGT	CATATG	TATTCGTGTTACAGCAA	GCGT	CTCGAG	GAAAAATCTTTAGGGAG
CP7102P	GTGCGT	CATATG	GCCGCTAAAGCAAAAT	GCGT	CTCGAG	TGAAAATCAAAGGTAGGT
CP7105P	GTGCGT	GCTAGC	AGTCTATATCAAAAATGGTG	GCGT	CTCGAG	ATCTTTCAATTTGGTTATCT
CP7106P	GTGCGT	CATATG	AAAGATTGGGGACTCT	GCGT	CTCGAG	GAATCCTAAGGCATACCTA
CP7107P	GTGCGT	GCTAGC	AGTATAGTCAGAAATTCGCA	GCGT	CTCGAG	GAAGCTAAGATTATAGCTACTTT
CP7108P	GTGCGT	GCTAGC	GCGGCCCTTTCCA	ACTCGCTA	GCGGCCGC	TTTATGTATATGGAACAGATAGG
CP7109P	GTGCGT	CATATG	GGACATTTTATTGATATTG	ACTCGCTA	GCGGCCGC	ATCATCAAGGTAGATAAAG
CP7110P	GTGCGT	CATATG	GGTTATTGCTATGTAATTACA	GCGT	CTCGAG	TTCTGATTGGACTCCA
CP7127P	GTGCGT	CATATG	GTGGCTTAACGATAGC	ACTCGCTA	GCGGCCGC	GCAGCCATCGTATTC
CP7130P	GTGCGT	CATATG	TTCAATATGCGAGG	GCGT	CTCGAG	CTTCTATTTTGAACPTTG
CP7140P	GTGCGT	CATATG	ACAGCCGGAGCAGCT	GCGT	CTCGAG	AGCACCCCTCAATTTTCATTG
CP7182P	GTGCGT	CATATG	GGATATGTTTTCTATGTGATC	GCGT	CTCGAG	GCTACTAAATCGAATCGA
CP6262P	GTGCGT	CATATG	ATCCCTGGATTAAAGTTCA	ACTCGCTA	GCGGCCGC	TTCACTGGGAGCTTGA
CP6268P	GTGCGT	CATATG	TACCAGGAGAATCTAAGAT	ACTCGCTA	GCGGCCGC	GATTTTCTTCTTCAGCTC
CP6296P	GTGCGT	CATATG	GAGGAGGTGTCTGAGTAT	ACTCGCTA	GCGGCCGC	ATGTTTCTTTTACTCTTTCT
CP6419P	GTGCGT	CATATG	GCTCCAGTCCGTGTT	GCGT	CTCGAG	AAGTGTTCTGTTGGAAGT
CP6601P	GTGCGT	CATATG	AATAAGCTACTCAATTTCTGT	GCGT	CTCGAG	GAAAATCTGAATTCCTTCCT
CP6639P	GTGCGT	CATATG	TTAAATTCAGCAATTC	GCGT	CTCGAG	AGGAACTAAACCTCATCT
CP6684P	GTGCGT	GCTAGC	GTTTTATTTCATGCTCAA	ACTCGCTA	GCGGCCGC	CTTAGAAAGACTATTTTCTAAGTA
CP6696P	GTGCGT	CATATG	TGCGTGATAATGGG	GCGT	CTCGAG	ATTCATCTTCGTAAAGAAT
CP6757P	GTGCGT	CATATG	GCAGTTGGTGGCGT	ACTCGCTA	GCGGCCGC	CTGTCCTCTGGAGC
CP6790P	GTGCGT	GCTAGC	AGTGAACACAAAAATCA	ACTCGCTA	GCGGCCGC	CTTATCGTCTGTATCAATA
CP6814P	GTGCGT	CATATG	CATGACGCACCTCTAAG	GCGT	CTCGAG	TACAGCTGCGCGA
CP6834P	GTGCGT	CATATG	GTTATGGGAACCTATATCG	GCGT	CTCGAG	TACATTTGTATTGATTTCAG
CP6878P	GTGCGT	CATATG	AACGTCCCTGATTCC	GCGT	CTCGAG	GCTAGCGGCTCTTTC
CP6892P	GTGCGT	CATATG	CAGAAGCATCCTTCCT	ACTCGCTA	GCGGCCGC	TCCTCTTTAGGAAATGG
CP6909P	GTGCGT	CATATG	TCCTCTTTAGGAAATGG	GCGT	CTCGAG	CAGTGCCAAAGTAGGGA
CP7015P	GTGCGT	CATATG	GCAGTACGATTAAATGTTG	GCGT	CTCGAG	TTTATTGTAGTCTATTTTATATTT
CP7035P	GTGCGT	GCTAGC	AGCAGAAAGACAATGA	GCGT	CTCGAG	ATTTTGAGTGTCTTGCA
CP7073P	GTGCGT	CATATG	ATTACCATAAATCACGTG	GCGT	CTCGAG	TATCCATCGACTTATAGC
CP7085P	GTGCGT	GCTAGC	TGTATTTTCCCTTACGTA	ACTCGCTA	GCGGCCGC	GGATTCTGCATACTCTG
CP7092P	GTGCGT	CATATG	TCTCCTCTTCTTAAAAAA	GCGT	CTCGAG	GGATTCATTACTGACCA
CP7093P	GTGCGT	CATATG	AAATACCGCTTCACG	GCGT	CTCGAG	ATTCTGTAGGGCTACGT
CP7094P	GTGCGT	CATATG	GTACACTTCTCTATAACCC	GCGT	CTCGAG	TAAGTTTGTATTGCGGTAT
CP7132P	GTGCGT	CATATG	TTGTTATTAGGGACTTTAGGA	GCGT	CTCGAG	TTTCCCAACCGCA
CP7133P	GTGCGT	CATATG	GCTGCGAATGCTC	GCGT	CTCGAG	TAATTTAATACTCTTTGAAGG
CP7177P	GTGCGT	CATATG	CCTACTCAAGTTAAAAACAGA	GCGT	CTCGAG	AAGTTTATATTTTACGACTT
CP7184P	GTGCGT	GCTAGC	CATATAGGATTTTGCCA	GCGT	CTCGAG	GTACTTAGCAAAGCGAT
CP7206P	GTGCGT	GCTAGC	AAGAAGCTATATCACCCCTA	GCGT	CTCGAG	CACACCGAGGAAAC
CP7222P	GTGCGT	CATATG	GTAGTTTCAGAAAGAAAGTC	GCGT	CTCGAG	ACGTATGCGCAACTG
CP7223P	GTGCGT	CATATG	GAAGTATTAGACCGCTCT	GCGT	CTCGAG	CGAGAAAAAGCTTCC
CP7224P	GTGCGT	CATATG	ATGAAGAAAAATTCGAAA	ACTCGCTA	GCGGCCGC	TAAGCATTCACAAATGA
CP7225P	GTGCGT	CATATG	CATATTTTGGCTTGATCGT	GCGT	CTCGAG	TCTTTTAACTAAATCTTGTTCTT
CP7303P	GTGCGT	CATATG	CTGTCTATTGTTTGTATCC	GCGT	CTCGAG	AAAATATACGGAACTCGC
CP7304P	GTGCGT	GCTAGC	GAAAGTTATAGTTTTCCTC	GCGT	CTCGAG	TTTTTGATTCTTAAAGAAG
CP7305P	GTGCGT	CATATG	GAAAGTTTATAGTTTTCACCTT	GCGT	CTCGAG	ACTCCTTGAGAAGGGAA
CP7307P	GTGCGT	CATATG	CTTAATCATGCTAAAAAGC	ACTCGCTA	GCGGCCGC	CTCTTTTATTTTAGGAAGCT

## CLAIMS

1. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 97,  
1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53,  
55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 99, 101, 103, 105,  
5 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143,  
145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181,  
183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219,  
221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257,  
259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293, 295,  
10 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333,  
335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371,  
373, 375, & 377.
2. A protein having 50% or greater sequence identity to a protein according to claim 1.
3. A protein comprising a fragment of an amino acid sequence selected from the group consisting of  
15 SEQ IDs 97, 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47,  
49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 99,  
101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137,  
139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175,  
177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213,  
20 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251,  
253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289,  
291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327,  
329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365,  
367, 369, 371, 373, 375, & 377.
- 25 4. A nucleic acid molecule which encodes a protein according to any one of claims 1 to 3.
5. A nucleic acid molecule according to claim 4, comprising a nucleotide sequence selected from  
the group consisting of SEQ IDs 98, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34,  
36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86,  
88, 90, 92, 94, 96, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128,  
30 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166,  
168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204,  
206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242,  
244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280,  
282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318,

6830	177.4	174; 91; 13	GST
6849	57.3	multiple bands	GST
6850	7.4-9.4	61; 14; 8	GST
6854	42.2	-	GST
6878	40.4	-	GST
6900	28.0	-	GST
6960	25.6	75; 35	GST
6968	34.6	83; 53; 35	GST
6998	39.3	multiple bands	GST
7033	68.2	multiple bands	GST
7101	113	105	GST
7102	63.4	-	GST
7105	29.2	30	GST
7106	39.5	72;46	GST
7107	71.4	67; 31	His
7108	35.9	35	GST
7111	46.1	51	GST
7132	17.9	57; 47; 17	His
7140	36.2-29.8	50; 38; 34	GST
7170	34.4	77; 33	GST
7224	39.4	40	GST
7287	167.3	180	GST
7306	50.1	50	GST

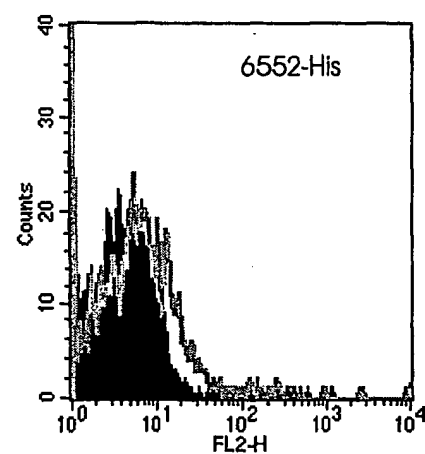
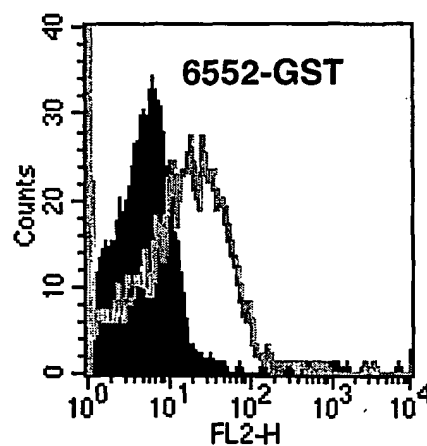
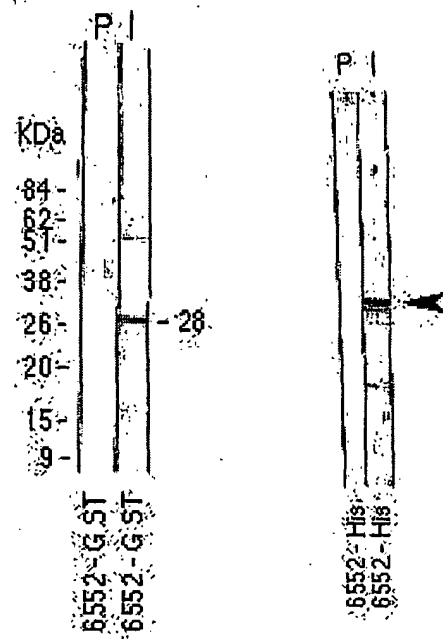
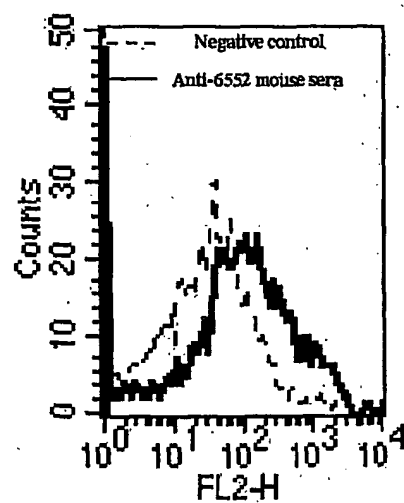
TABLE IV – FACS-positive proteins not found in *C.trachomatis*

cp7105	cp6390
cp7106	cp6784
cp7107	cp6296
cp7108	

TABLE V – Proteins identified by MALDI-TOF following 2D electrophoresis

cp6270	cp6733	cp6900
cp6552	cp6736	cp6960
cp6576	cp6737	cp6998
cp6577	cp6752	cp7033
cp6602	cp6767	cp7108
cp6664	cp6784	cp7111
cp6727	cp6790	cp7170
cp6728	cp6830	cp7287
cp6729	cp6849	cp7306

1/169

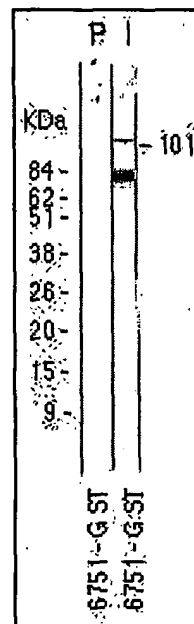
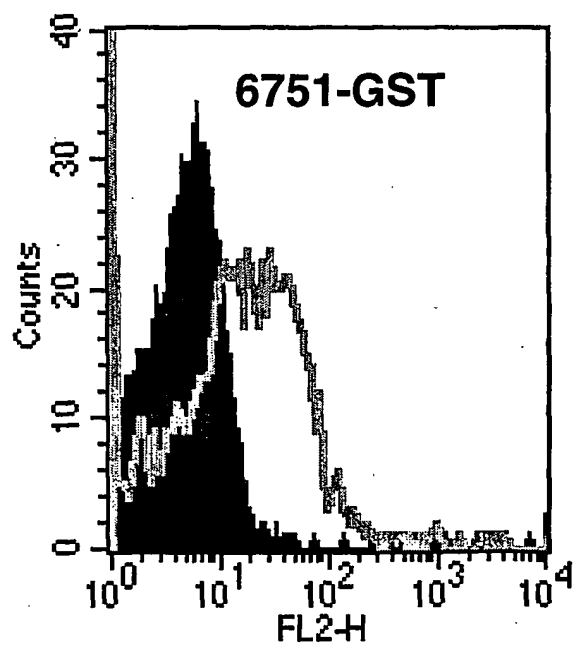
**FIGURE 1****Fig. 1A****Fig. 1B****Fig. 1C**



320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, & 378.

- 5 6. A nucleic acid molecule comprising a fragment of a nucleotide sequence selected from the group consisting of SEQ IDs 98, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 10 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, & 378.
- 15 7. A nucleic acid molecule comprising a nucleotide sequence complementary to a nucleic acid molecule according to any one of claims 4 to 6.
8. A nucleic acid molecule comprising a nucleotide sequences having 50% or greater sequence identity to a nucleic acid molecule according to any one of claims 4 to 7.
9. A nucleic acid molecule which can hybridise to a nucleic acid molecule according to any one of claims 4 to 8 under high stringency conditions.
- 20 10. A composition comprising a protein or a nucleic acid molecule according to any preceding claim.
11. A composition according to claim 10 being a vaccine composition.
12. A composition according to claim 10 or claim 11 for use as a pharmaceutical.
13. The use of a composition according to claim 10 in the manufacture of a medicament for the treatment or prevention of infection due to *Chlamydia* bacteria, particularly *Chlamydia* 25 *pneumoniae*.

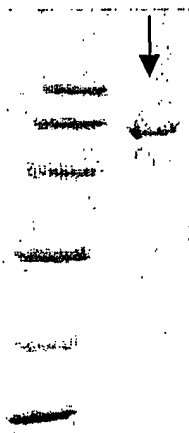
3/169

**FIGURE 3****FIG. 3A****FIG. 3B****FIG. 3C**

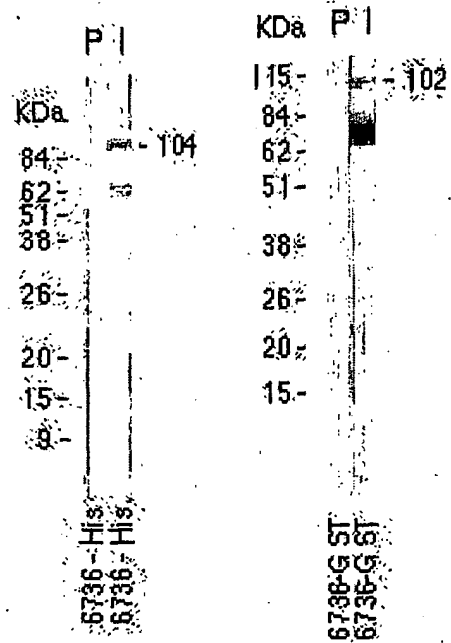
2/169

# **FIGURE 2**

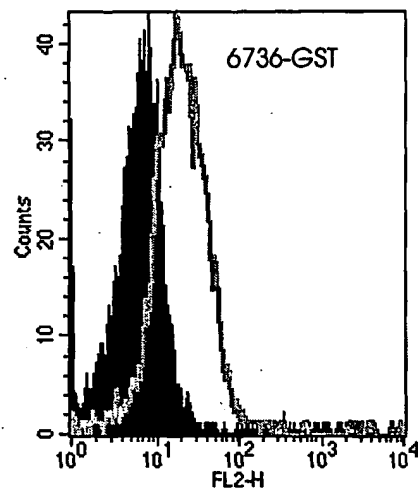
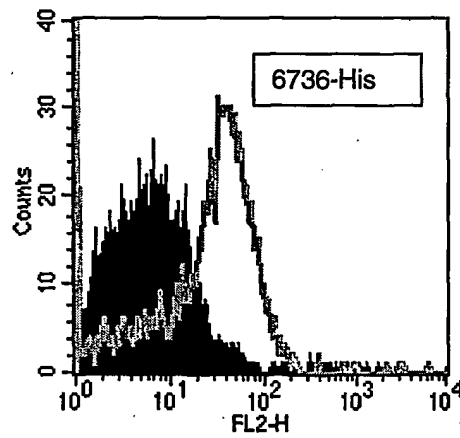
**Fig. 2A**



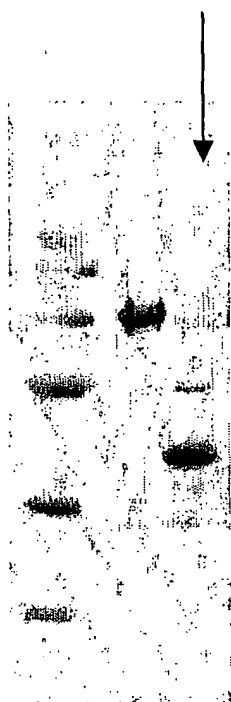
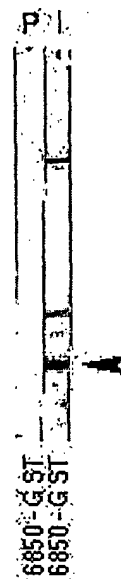
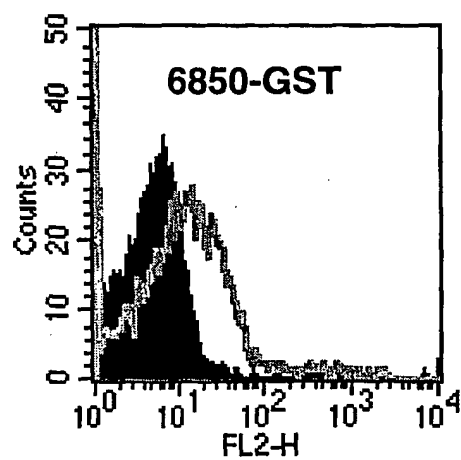
**Fig. 2B**



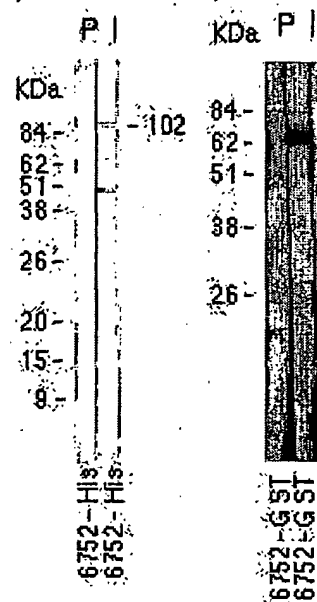
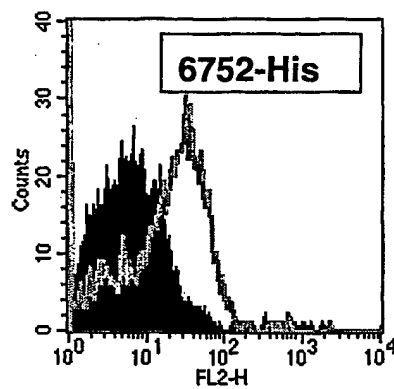
**Fig. 2C**



5/169

**FIGURE 5****Fig. 5A****Fig. 5B****Fig. 5C**

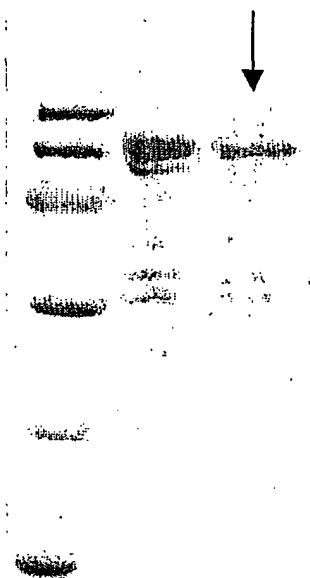
4/169

**FIGURE 4****FIG. 4A****FIG. 4B****FIG. 4C**

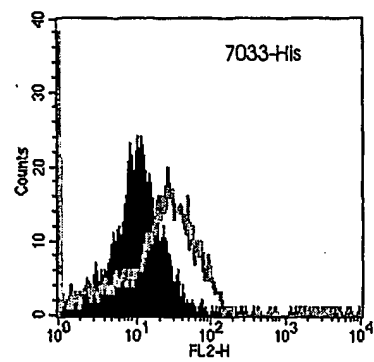
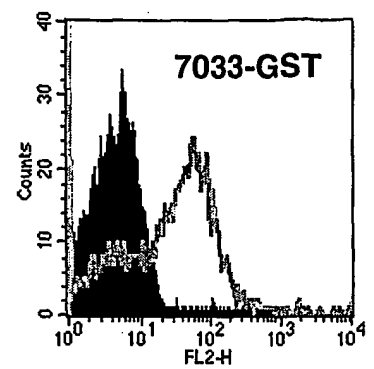
7/169

**FIGURE 7**

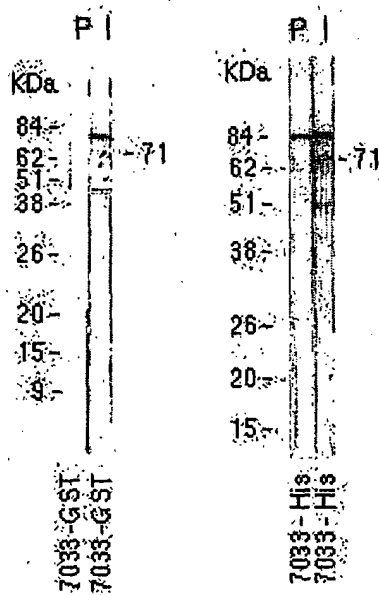
**Fig. 7A**



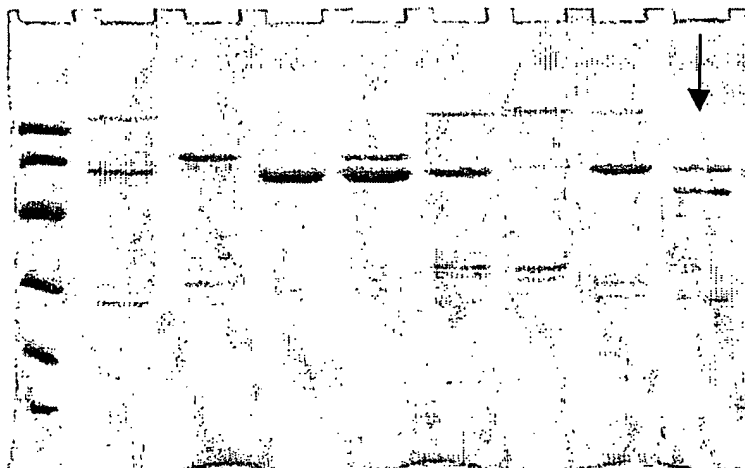
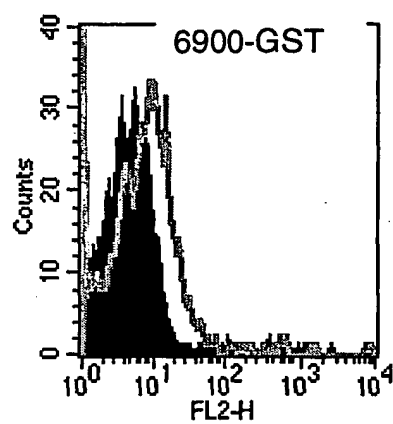
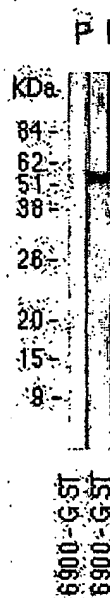
**Fig. 7B**



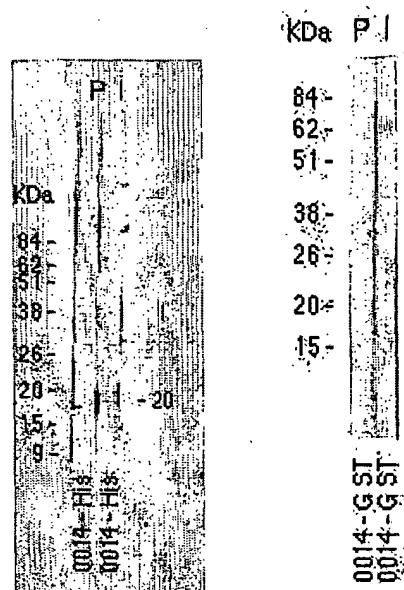
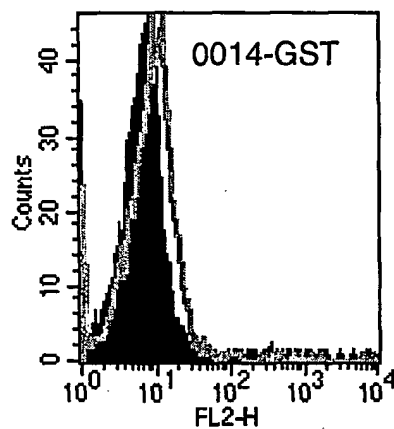
**Fig. 7c**



6/169

**FIGURE 6****Fig. 6A****Fig. 6B****Fig. 6C**

9/169

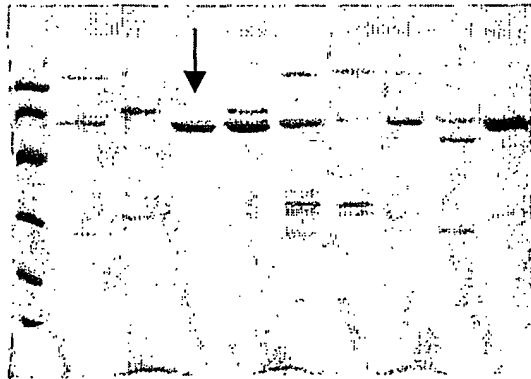
**FIGURE 9****FIG. 9A****FIG. 9B****FIG. 9C**



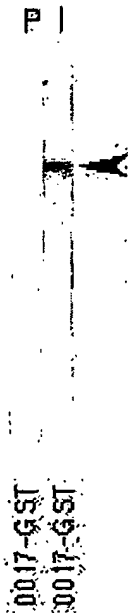
8/169

**FIGURE 8**

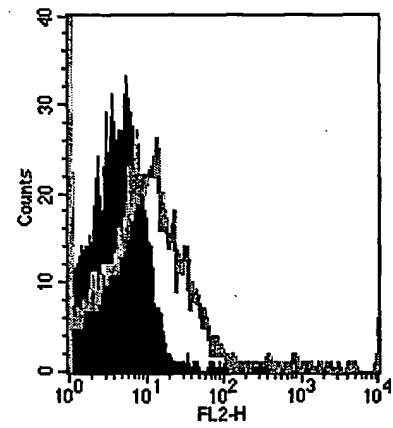
**FIG. 8A**



**FIG. 8B**



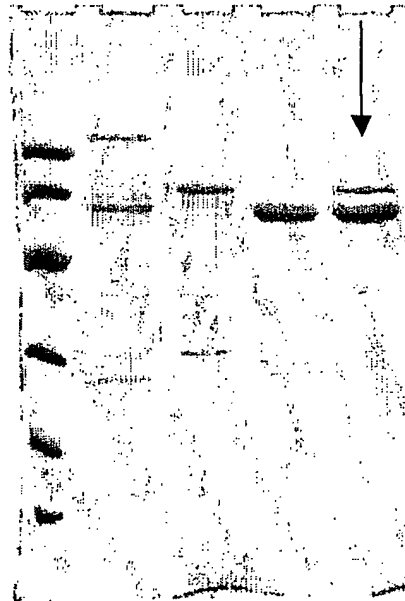
**FIG. 8C**



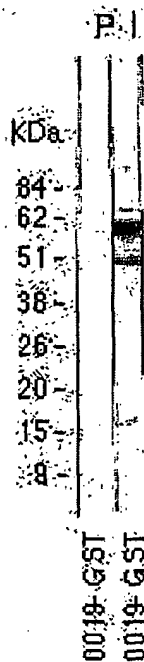
11/169

**FIGURE 11**

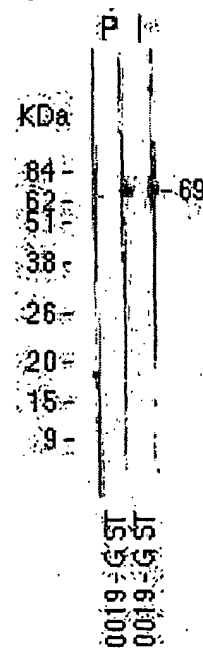
**FIG. 11A**



**FIG. 11B**



**FIG. 11C**



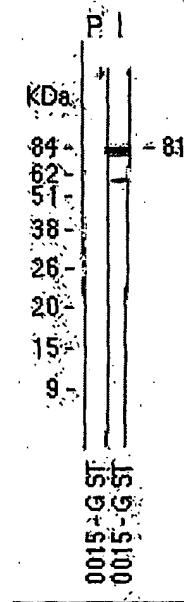
10/169

**FIGURE 10**

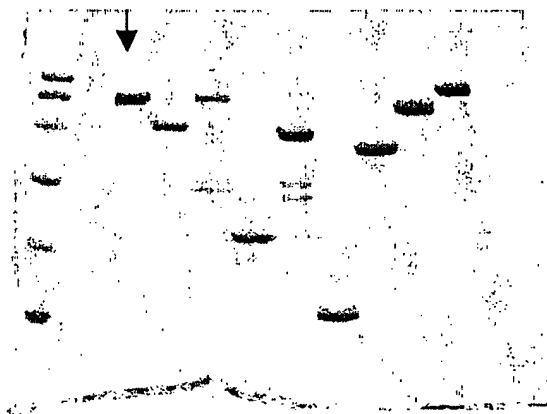
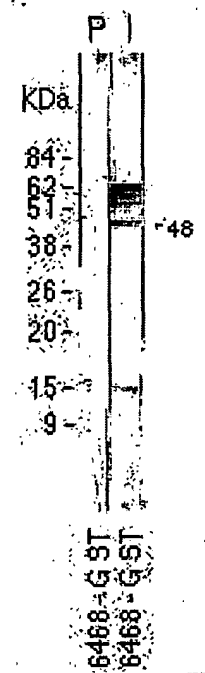
**FIG. 10A**



**FIG. 10B**



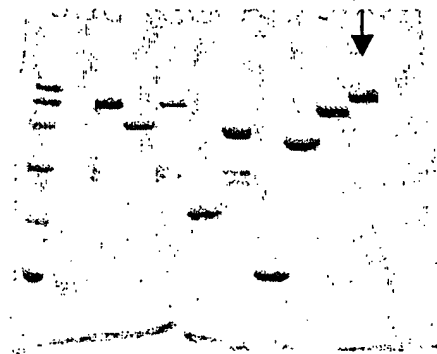
13/169

**FIGURE 13****FIG. 13A****FIG. 13B**

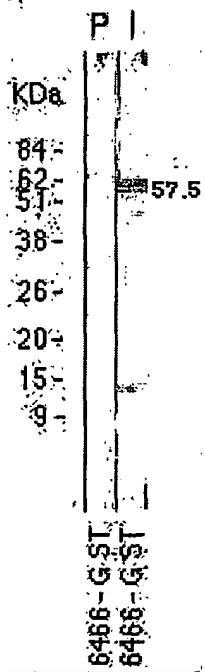
12/169

**FIGURE 12**

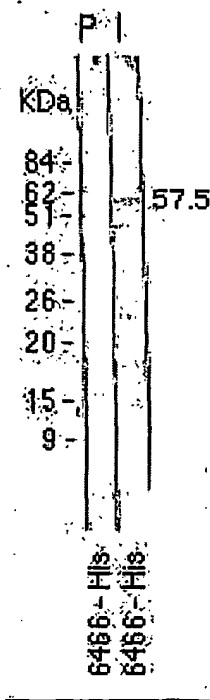
**Fig. 12A**



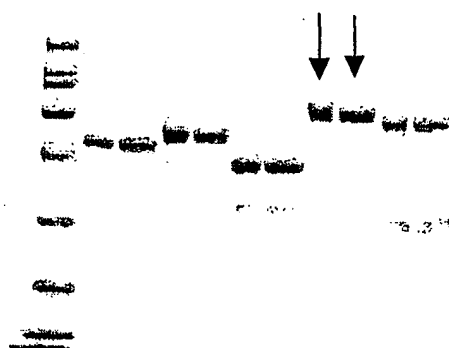
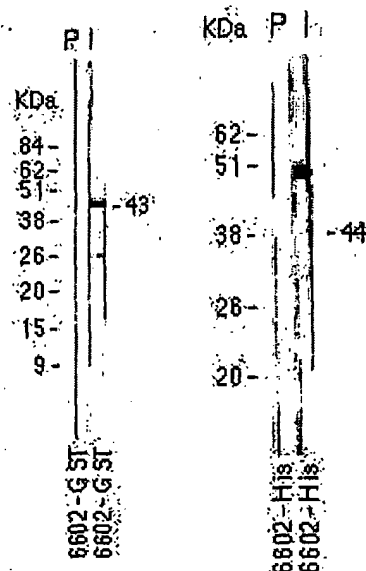
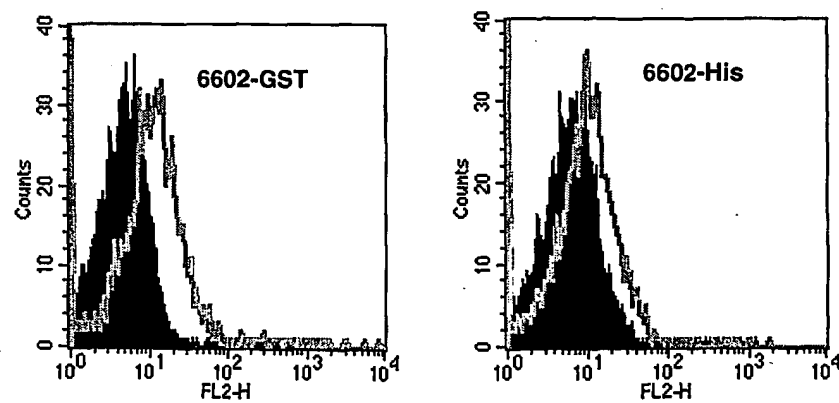
**Fig. 12B**



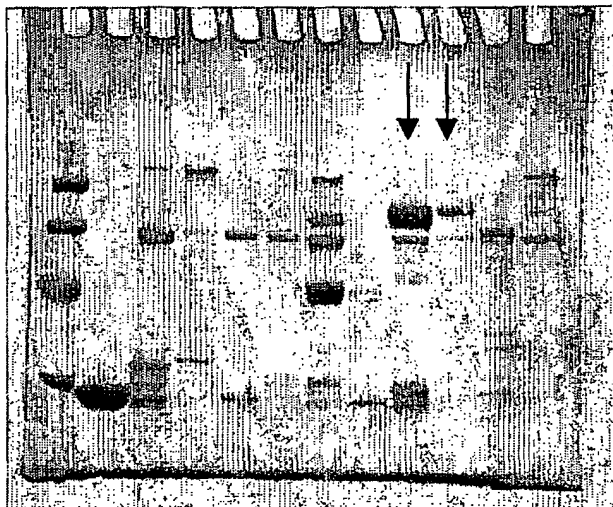
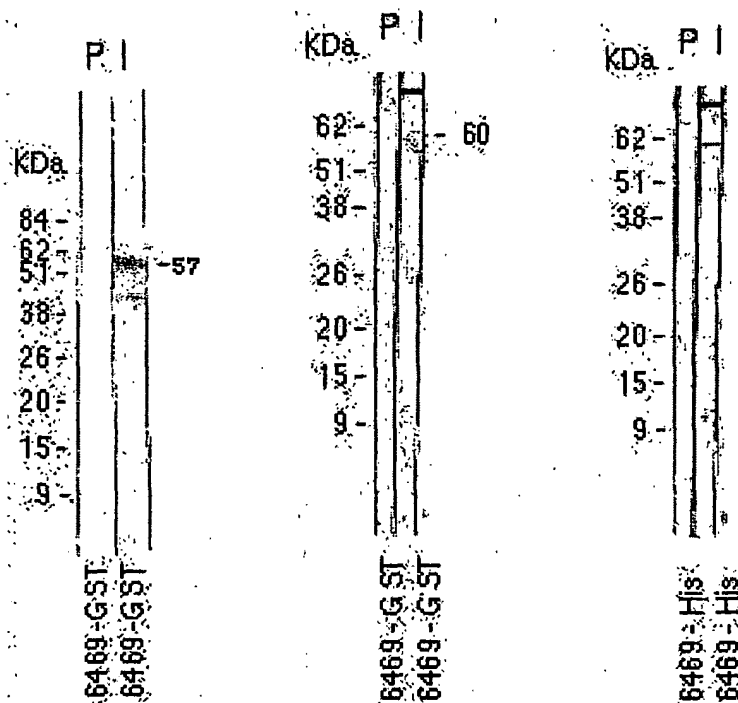
**Fig. 12C**



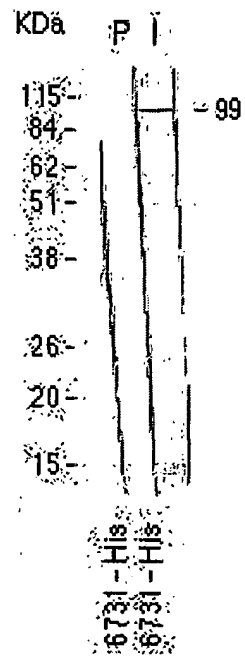
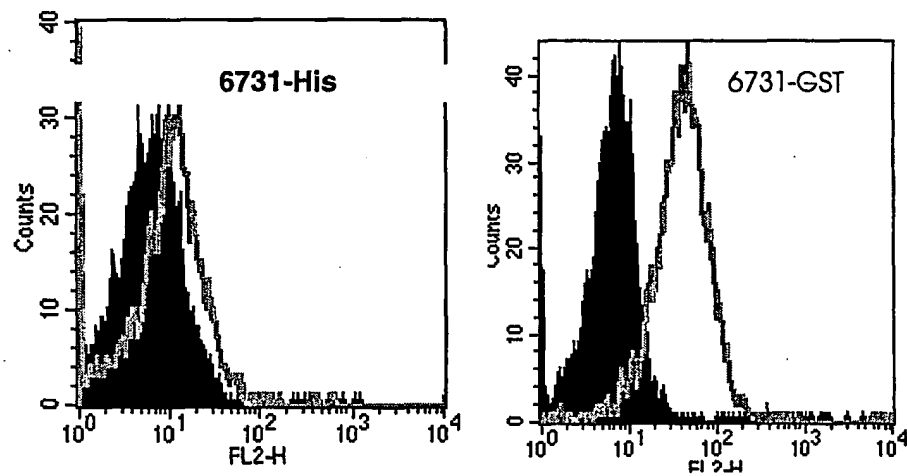
15/169

**FIGURE 15****FIG. 15A****Fig. 15B****FIG. 15C**

14/169

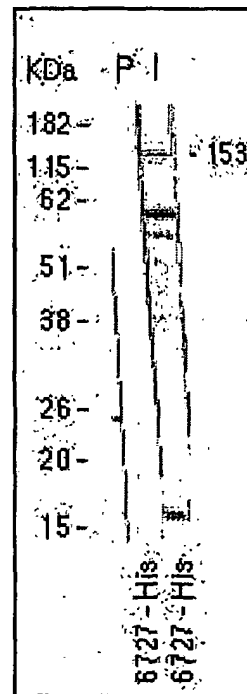
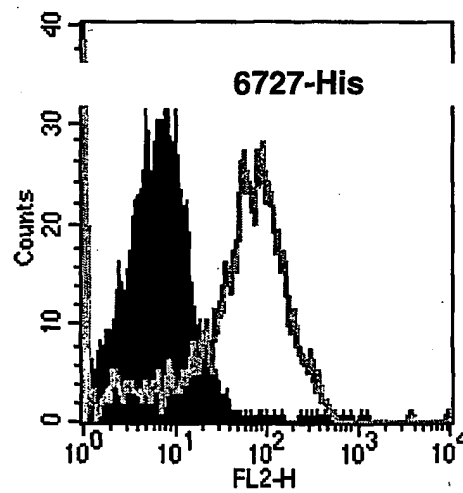
**FIGURE 14****Fig. 14A****Fig. 14B**

17/169

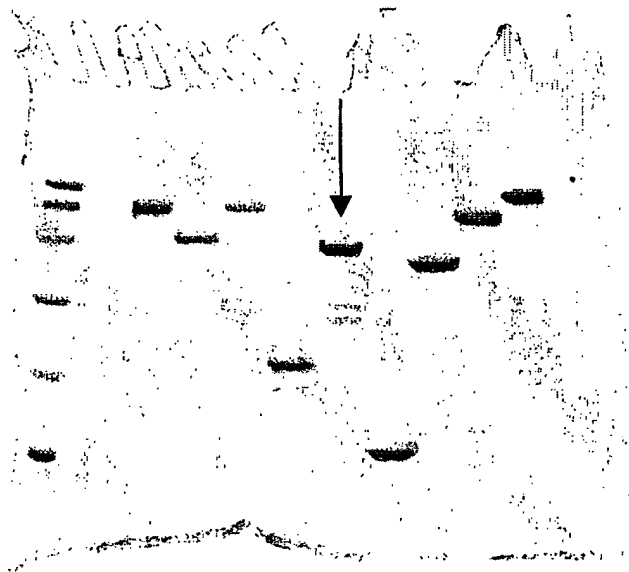
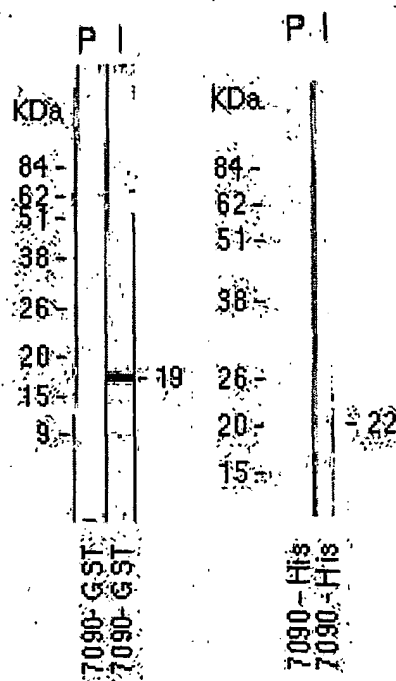
**FIGURE 17****Fig. 17A****Fig. 17B****Fig. 17C**



16/169

**FIGURE 16****Fig. 16A****Fig. 16B****Fig. 16C**

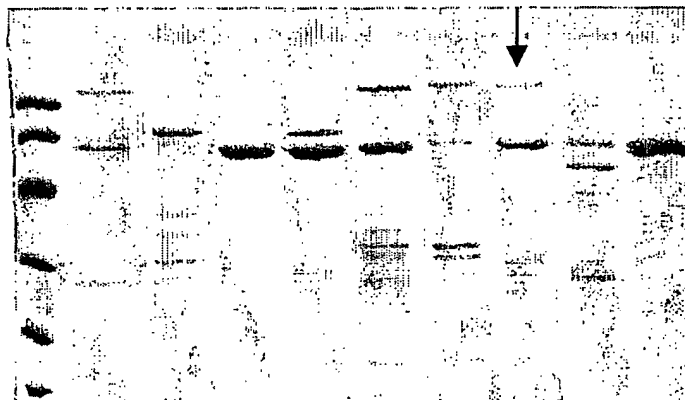
19/169

**FIGURE 19****FIG. 19A****FIG. 19B**

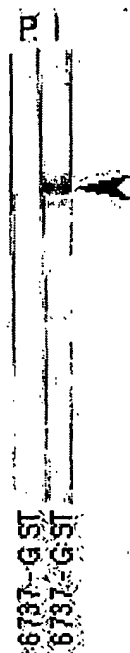
18/169

# **FIGURE 18**

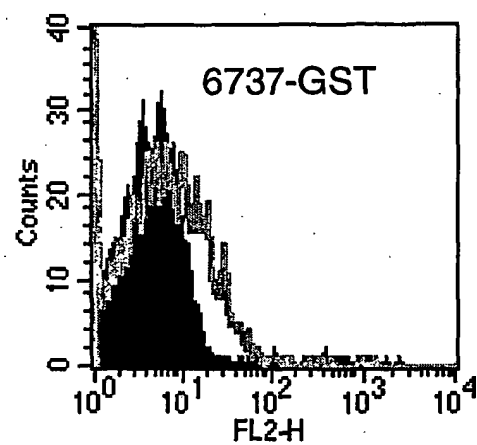
**Fig. 18A**



**Fig. 18B**



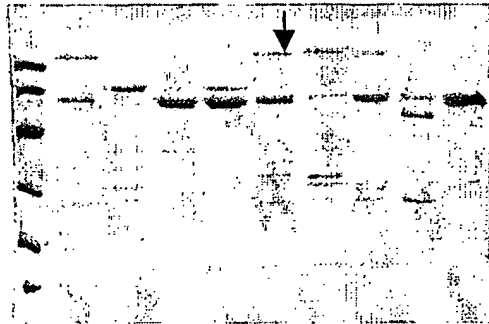
**Fig. 18C**



21/169

**FIGURE 21**

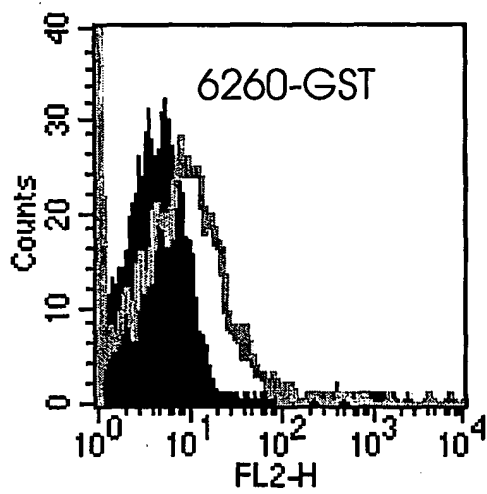
**FIG.  
21A**



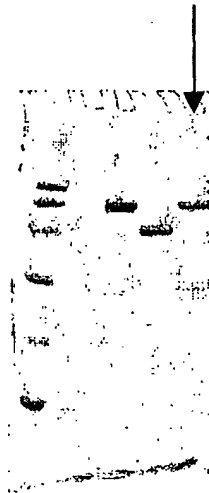
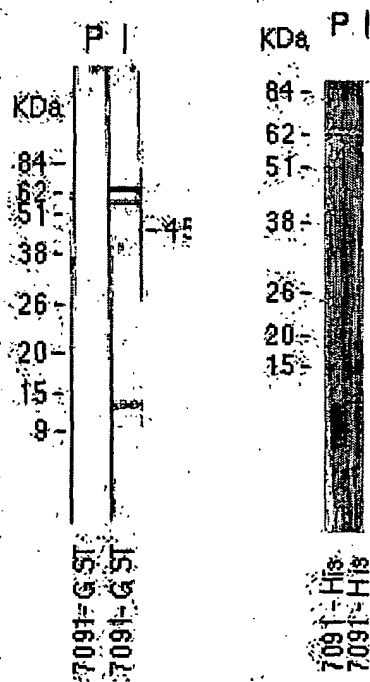
**FIG.  
21B**



**FIG.  
21C**



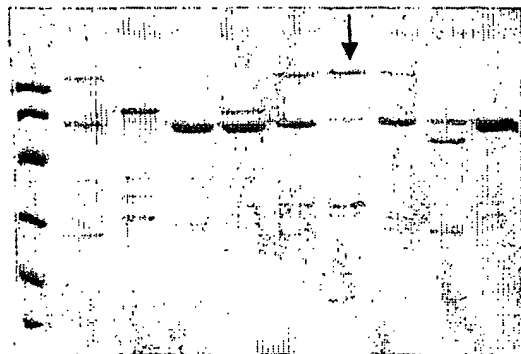
20/169

**FIGURE 20****FIG. 20A****FIG. 20B**

23/169

**FIGURE 23**

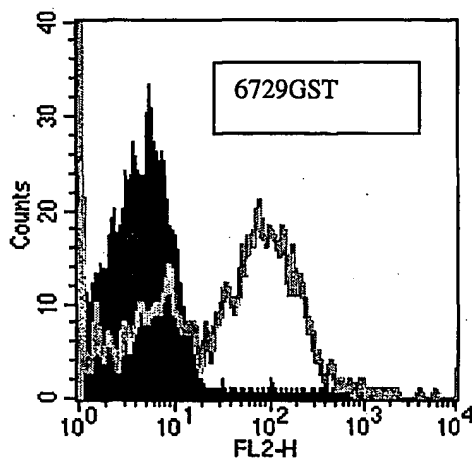
**FIG.  
23A**



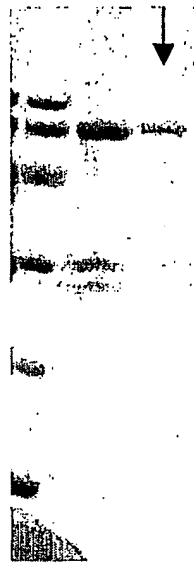
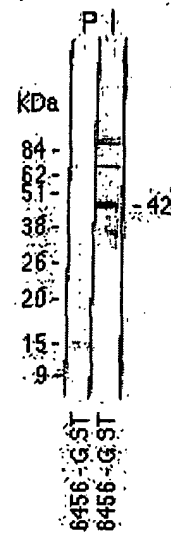
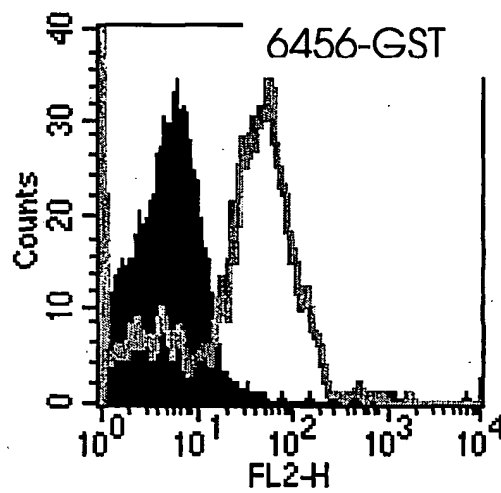
**FIG.  
23B**



**FIG.  
23C**



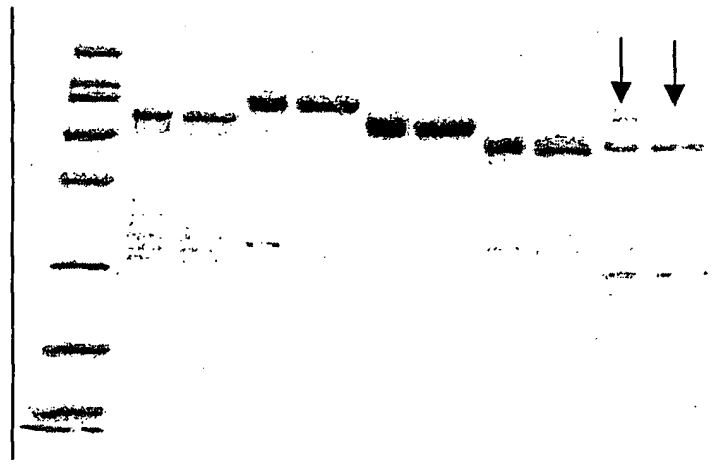
22/169

**FIGURE 22****FIG.  
22A****FIG.  
22B****FIG.  
22C**

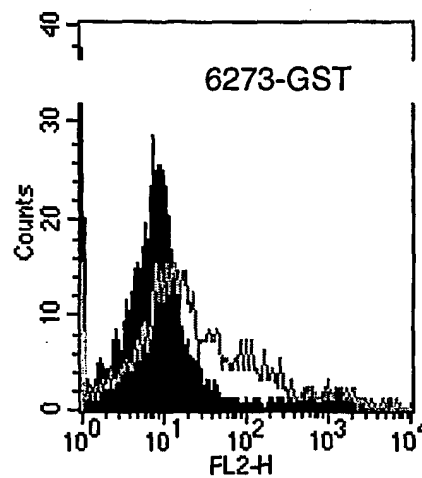
25/169

# **FIGURE 25**

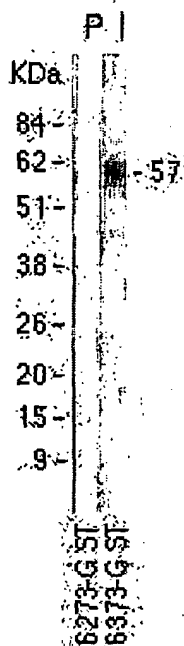
**Fig. 25A**



**Fig. 25C**

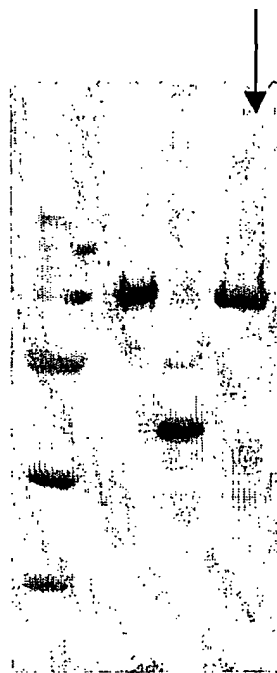
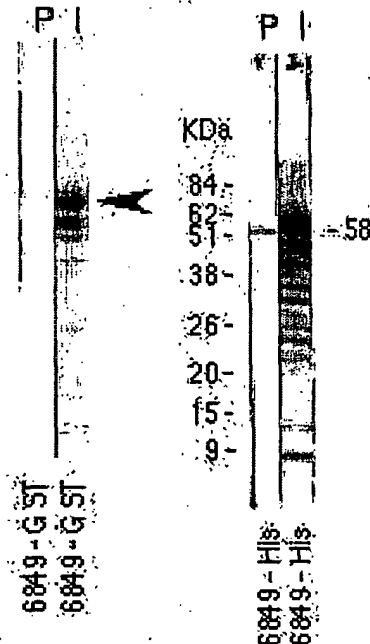
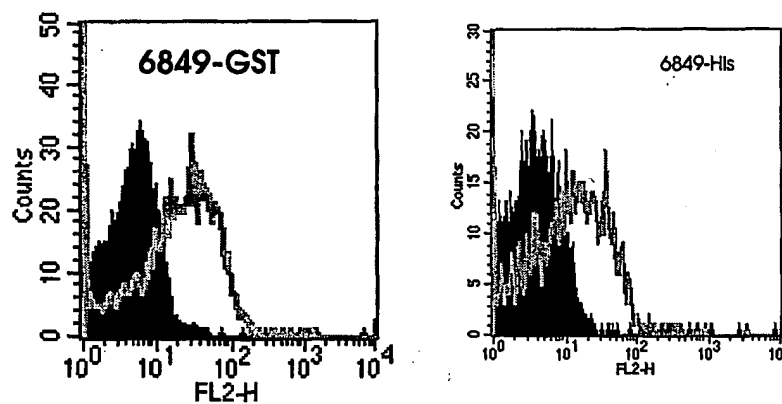


**Fig. 25B**



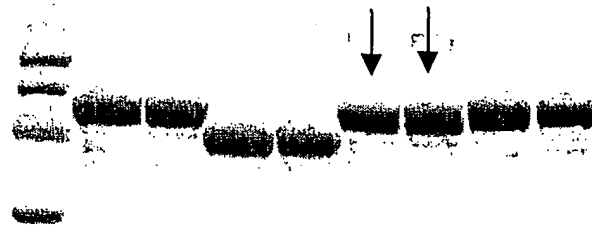


24/169

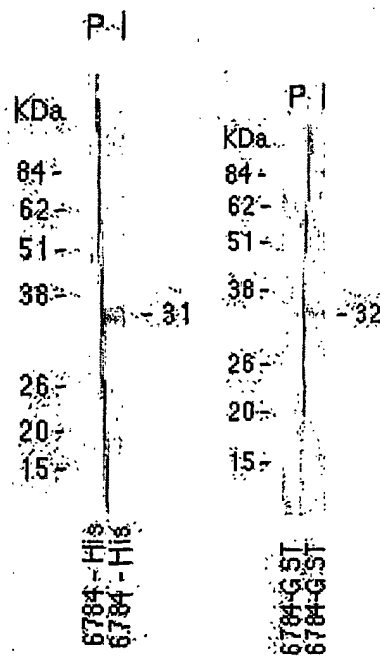
**FIGURE 24****FIG.  
24A****FIG.  
24B****FIG.  
24C**

27/169

**FIGURE 27**

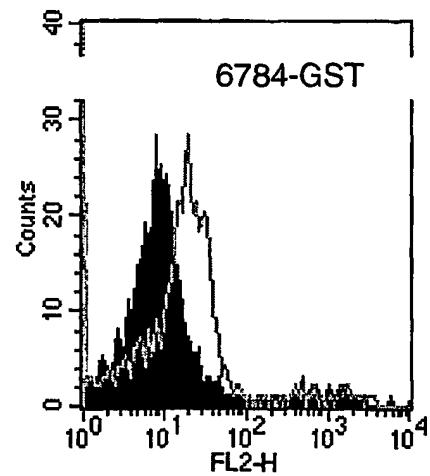


**Fig. 27A**

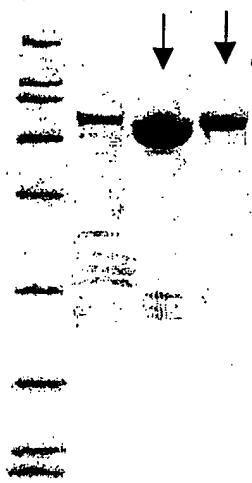
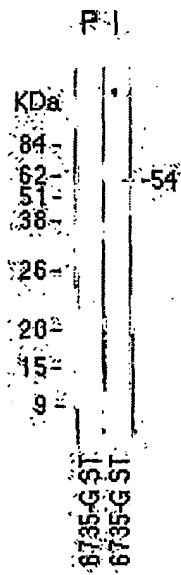


**Fig. 27B**

**Fig. 27C**



26/169

**FIGURE 26****FIG. 26A****FIG. 26B**

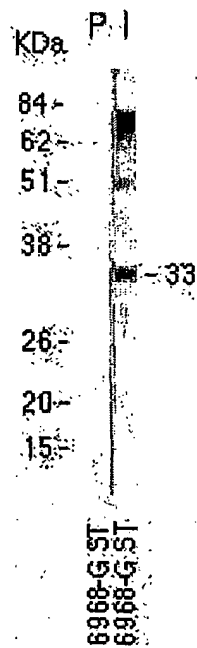
29/169

**FIGURE 29**

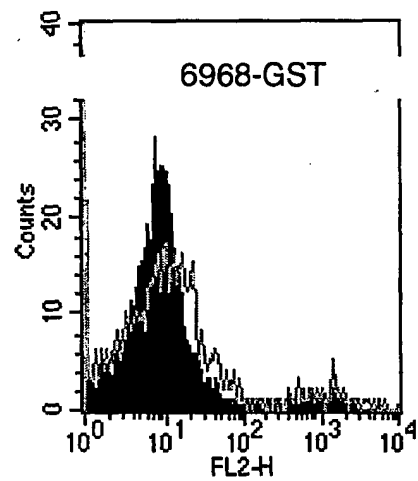
**Fig. 29A**



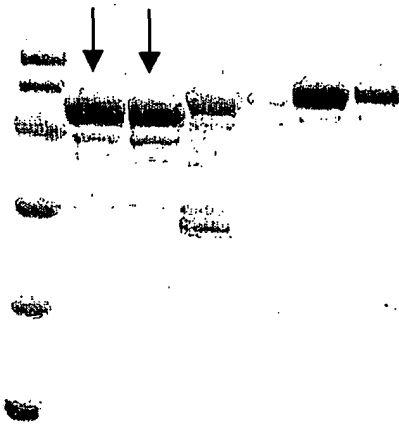
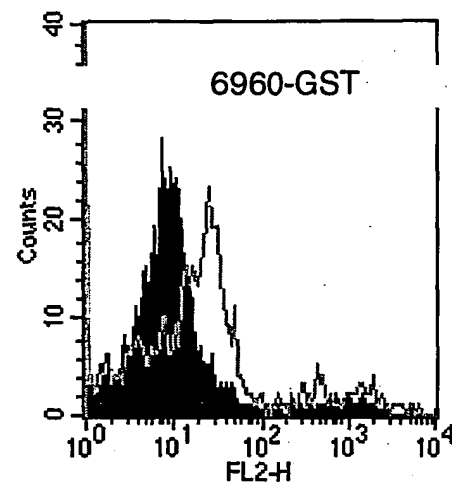
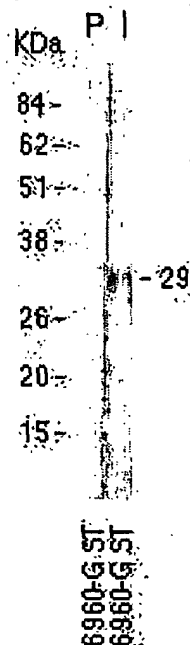
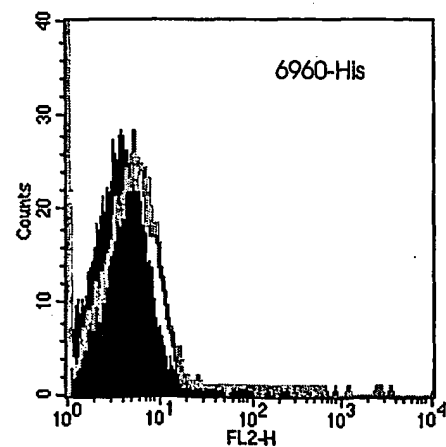
**Fig. 29B**



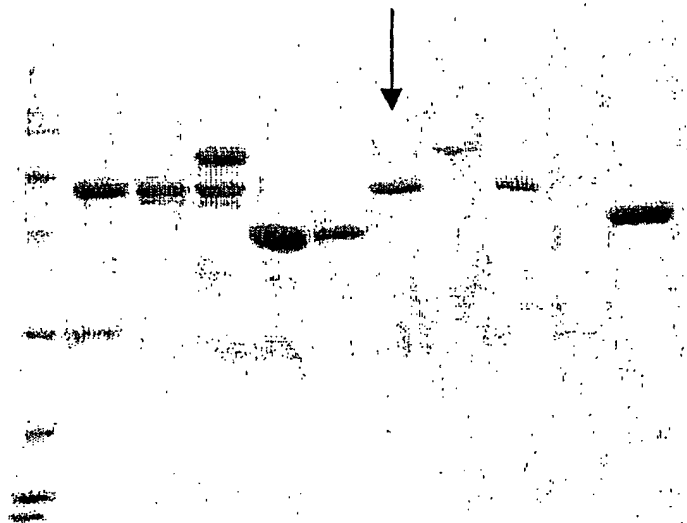
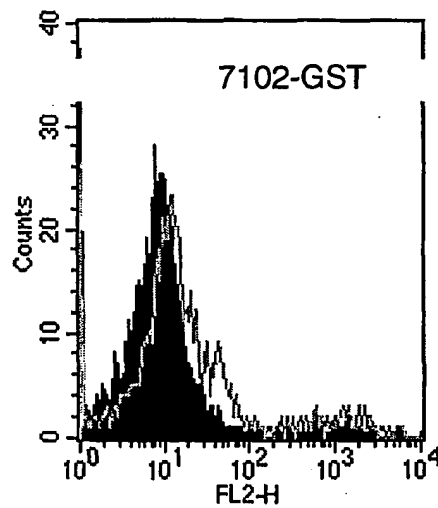
**Fig. 29C**



28/169

**FIGURE 28****FIG. 28A****FIG. 28B****FIG. 28C**

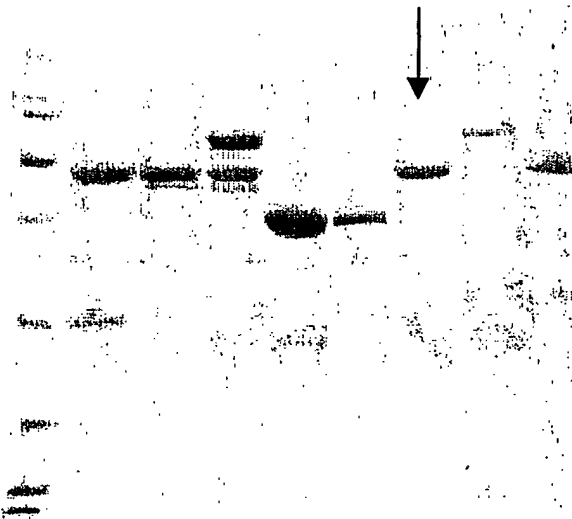
31/169

**FIGURE 31****Fig. 31A****Fig. 31B**

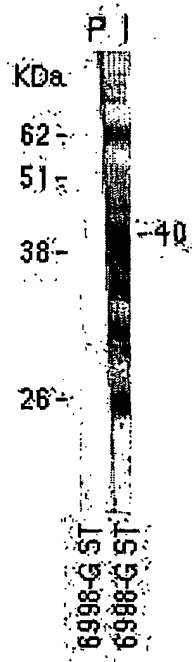
30/169

**FIGURE 30**

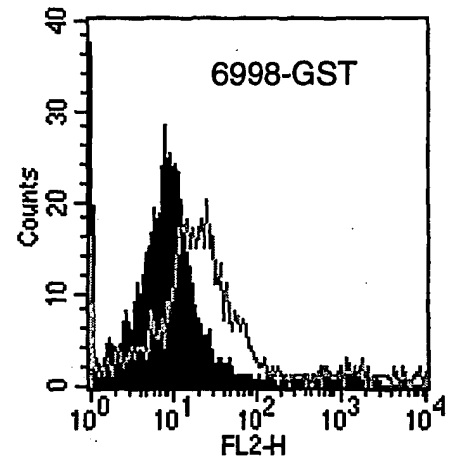
**Fig. 30A**



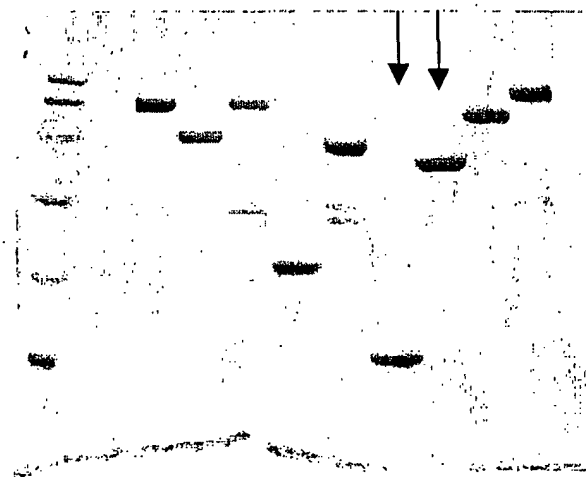
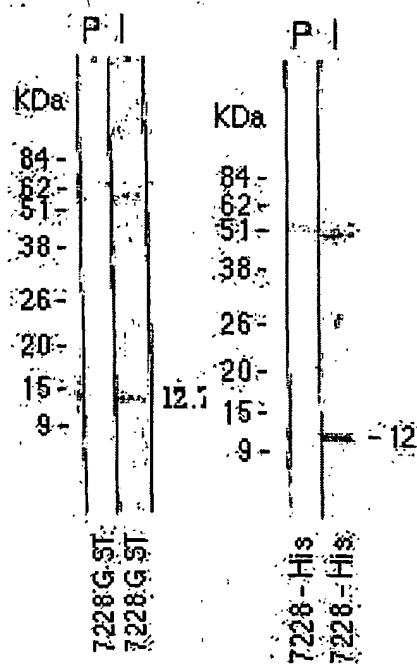
**Fig. 30B**



**Fig. 30C**

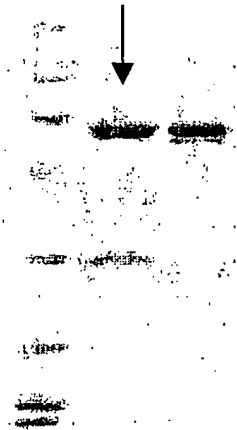
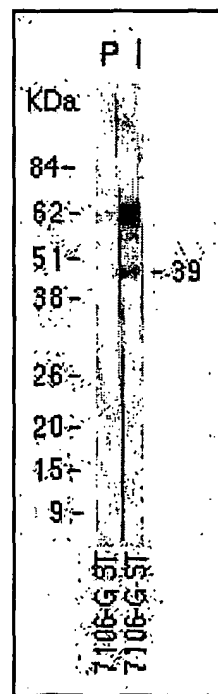
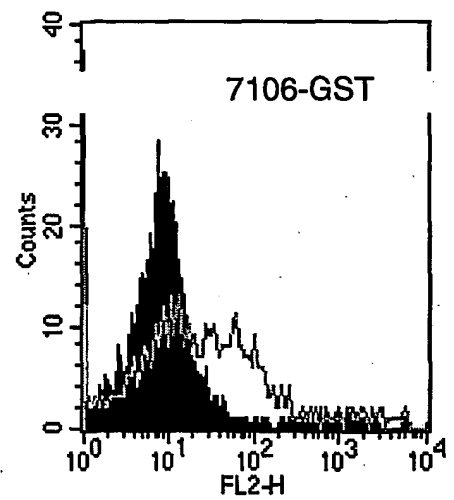


33/169

**FIGURE 33****FIG. 33A****FIG. 33B**



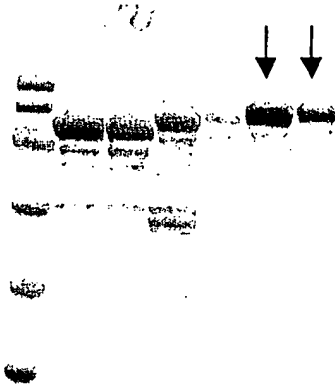
32/169

**FIGURE 32****Fig. 32A****Fig. 32B****Fig. 32C**

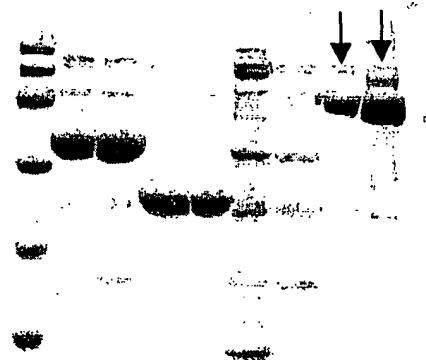
35/169

**FIGURE 35**

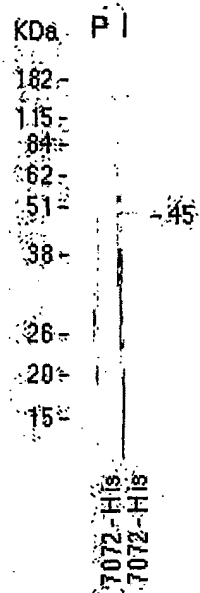
**FIG. 35A**



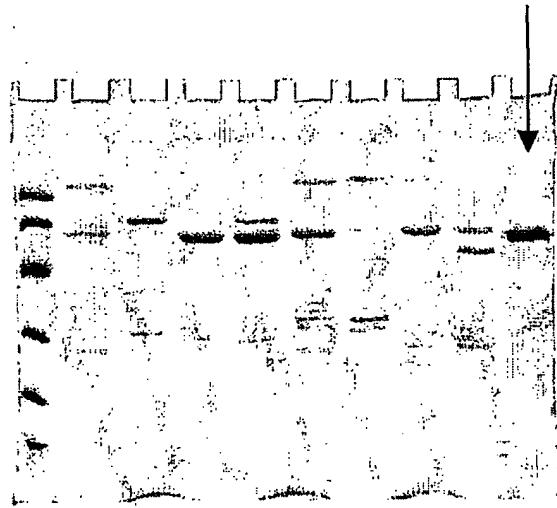
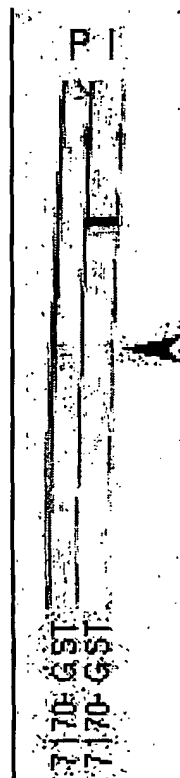
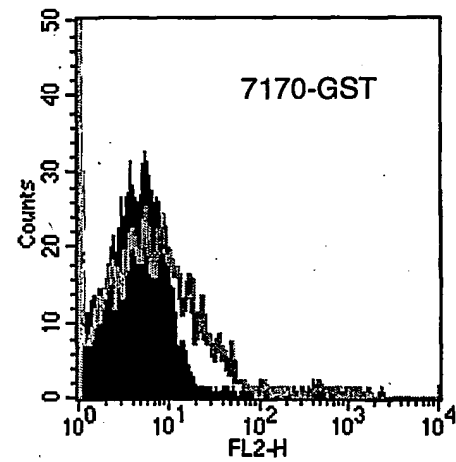
**FIG. 35B**



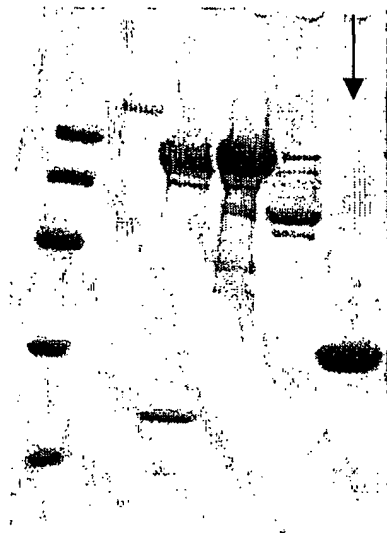
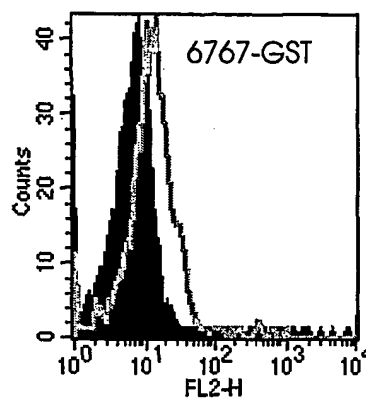
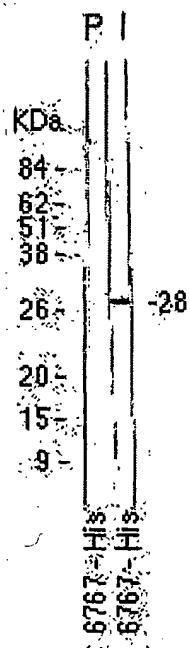
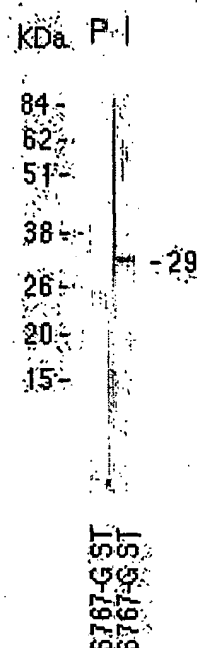
**FIG. 35C**



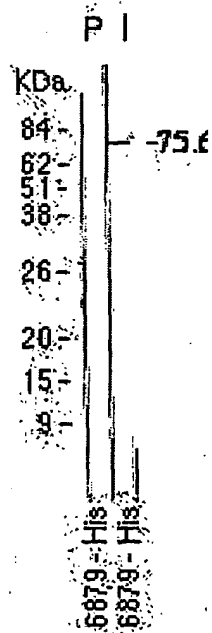
34/169

**FIGURE 34****FIG. 34A****FIG. 34B****FIG. 34C**

37/169

**FIGURE 37****FIG. 37A****FIG. 37C****FIG. 37B****FIG. 37D**

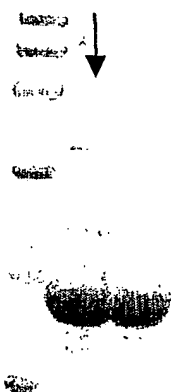
36/169

**FIGURE 36****Fig. 36A****Fig. 36B**

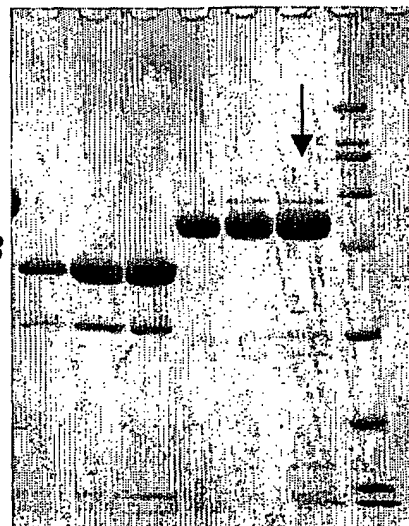
39/169

# **FIGURE 39**

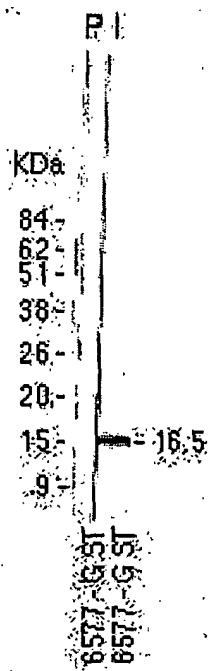
**FIG. 39A**



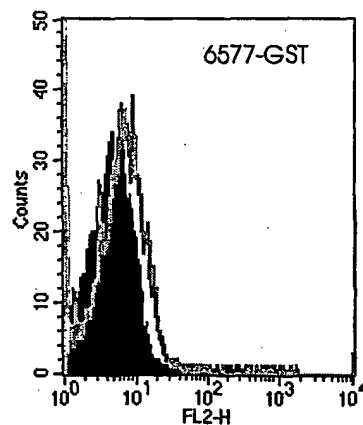
**FIG. 39B**



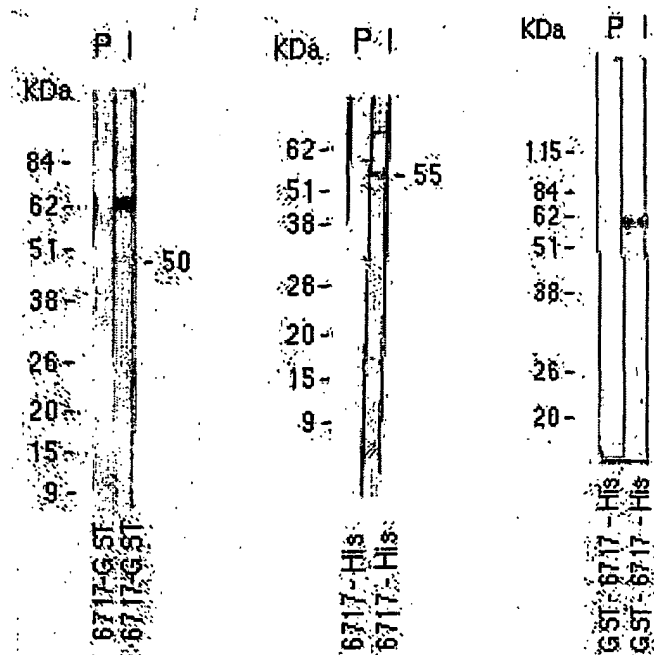
**FIG. 39C**



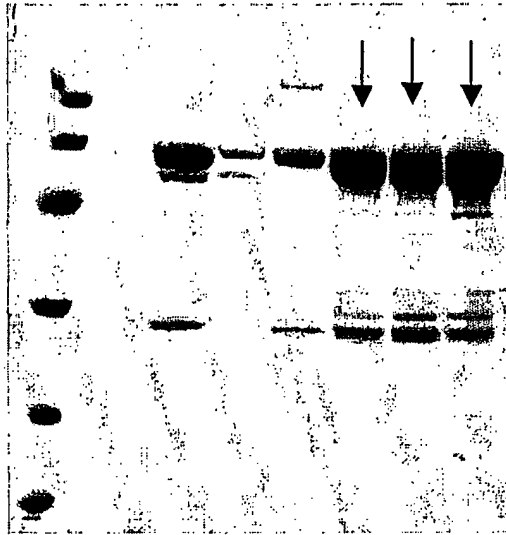
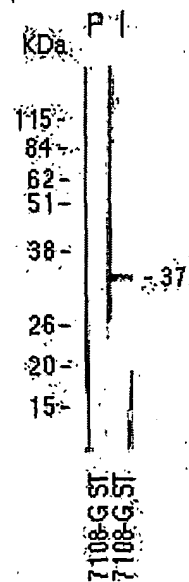
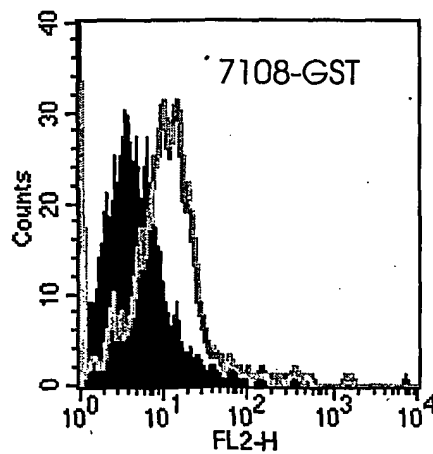
**FIG. 39D**



38/169

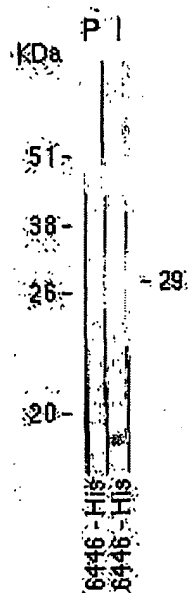
**FIGURE 38****Fig. 38A****Fig. 38B**

41/169

**FIGURE 41****FIG. 41A****FIG. 41B****FIG. 41C**



40/169

**FIGURE 40****FIG. 40A****FIG. 40B**

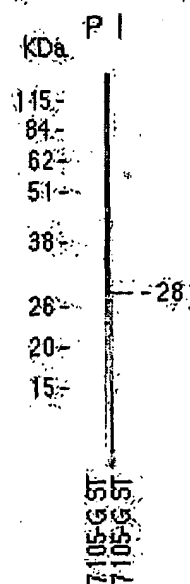
43/169

# **FIGURE 43**

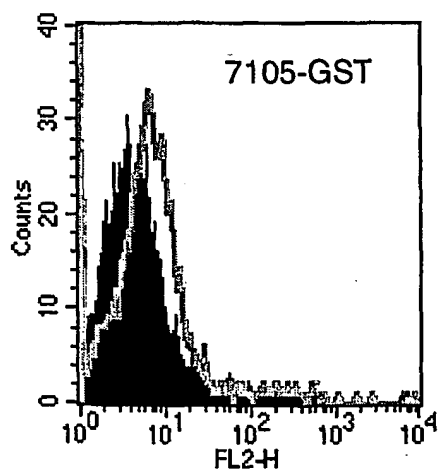
**Fig. 43A**



**Fig. 43B**



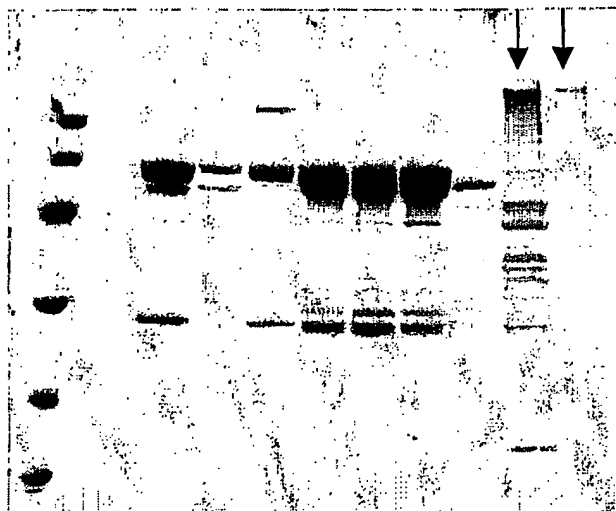
**Fig. 43C**



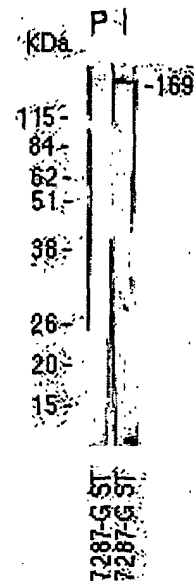
42/169

**FIGURE 42**

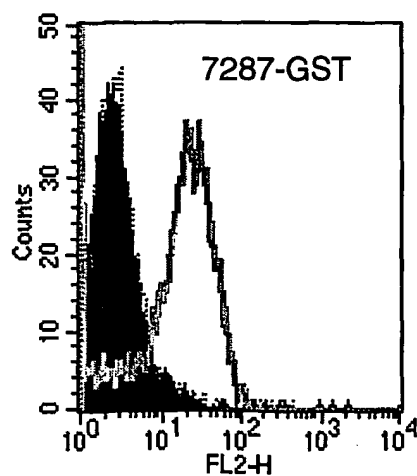
**FIG. 42A**



**Fig. 42B**



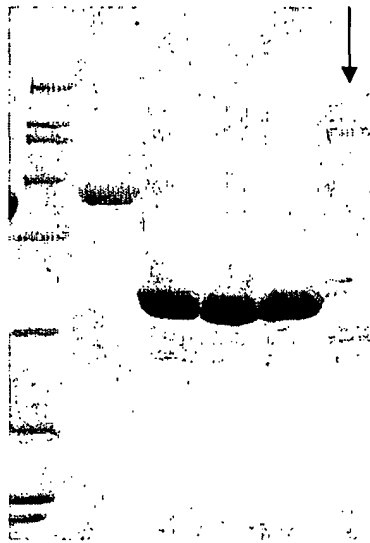
**FIG. 42C**



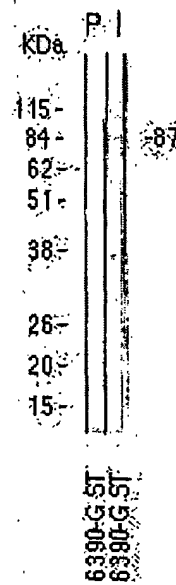
45/169

**FIGURE 45**

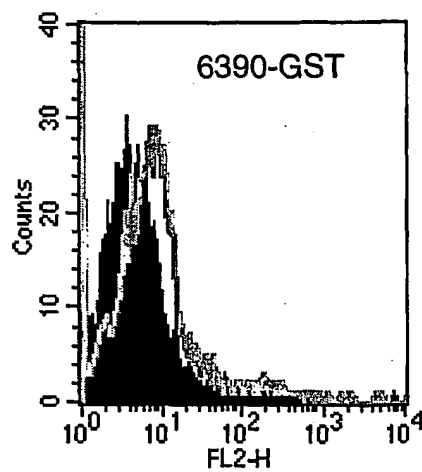
**Fig. 45A**



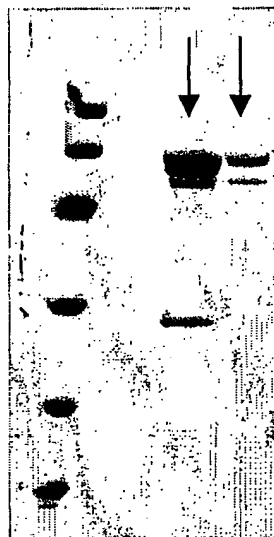
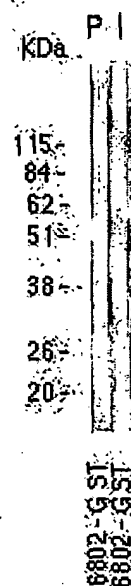
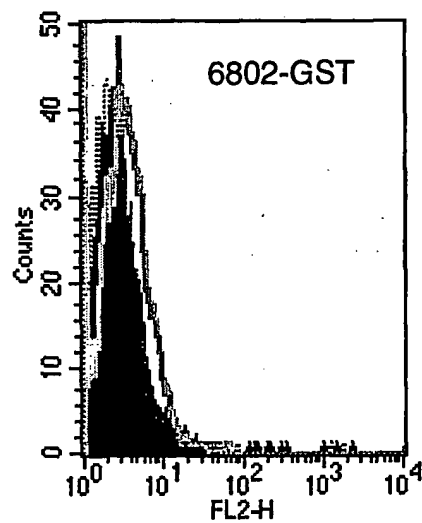
**Fig. 45B**



**Fig. 45C**



44/169

**FIGURE 44****FIG. 44A****FIG. 44B****FIG. 44C**

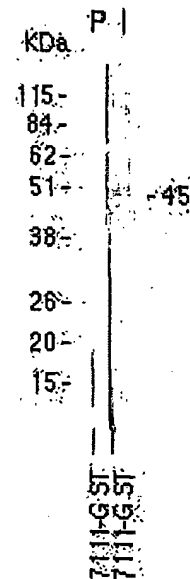
47/169

# **FIGURE 47**

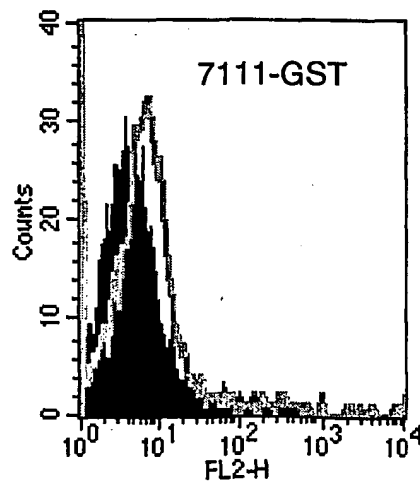
**Fig. 47A**



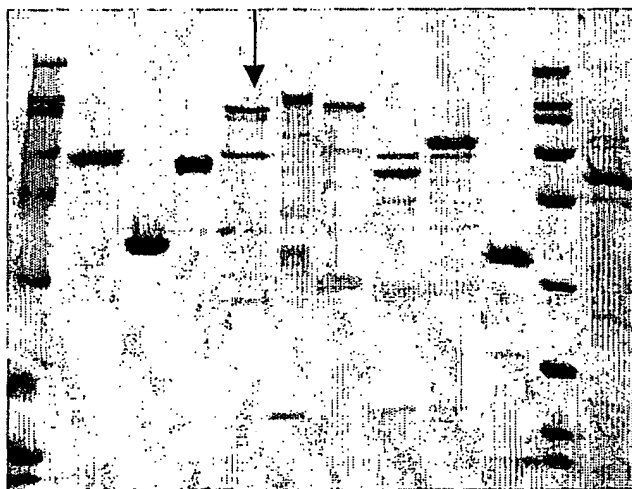
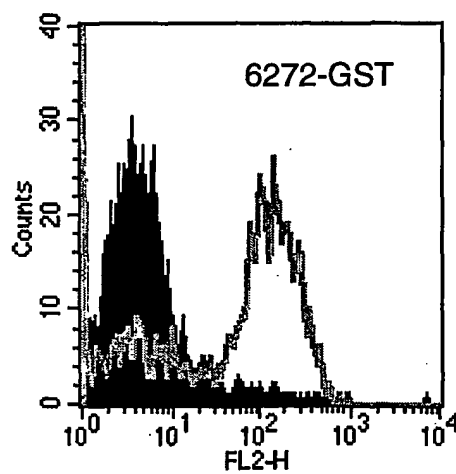
**FIG. 47B**



**Fig. 47C**



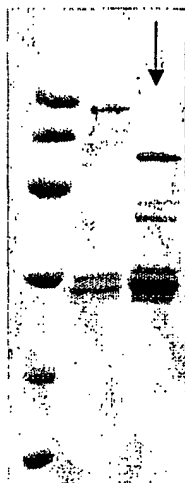
46/169

**FIGURE 46****Fig. 46A****Fig. 46B**

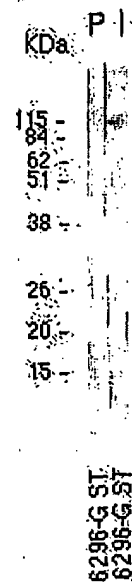
49/169

**FIGURE 49**

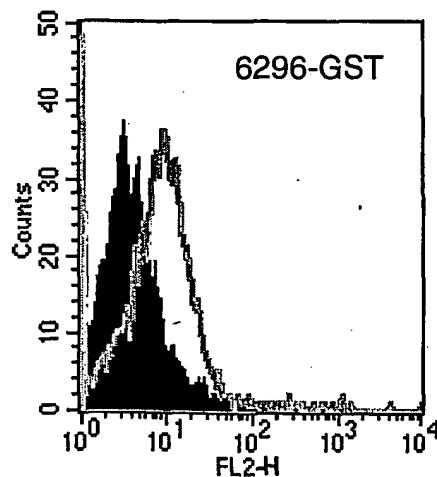
**Fig. 49A**



**FIG. 49B**

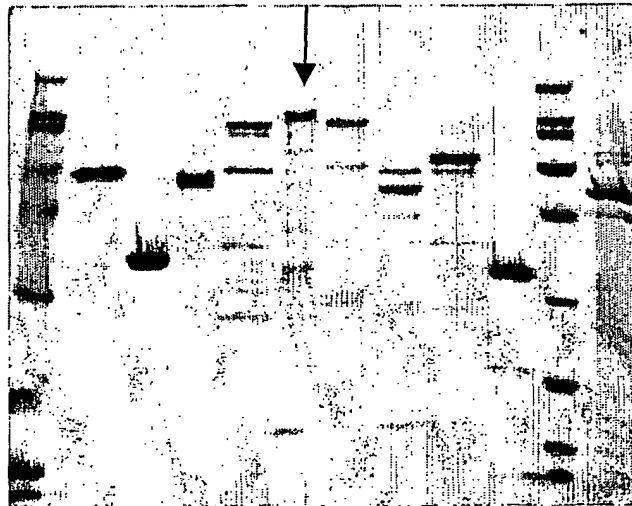
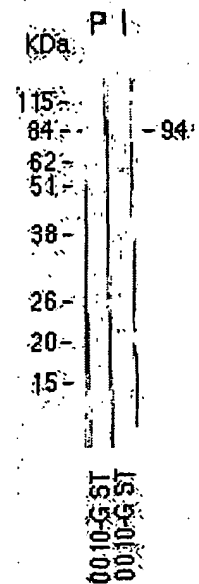
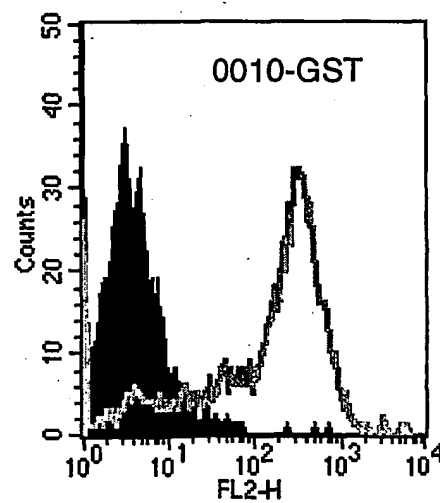


**FIG. 49C**





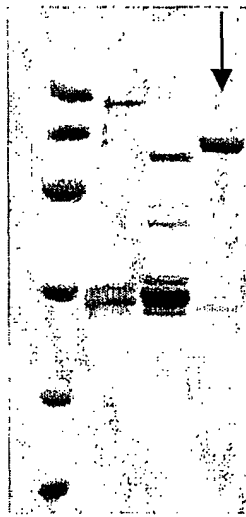
48/169

**FIGURE 48****FIG. 48A****FIG. 48B****FIG. 48C**

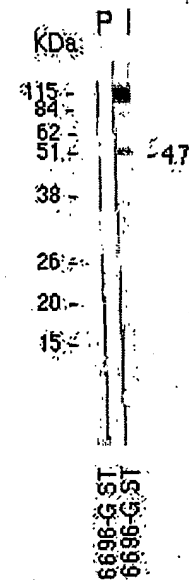
51/169

**FIGURE 51**

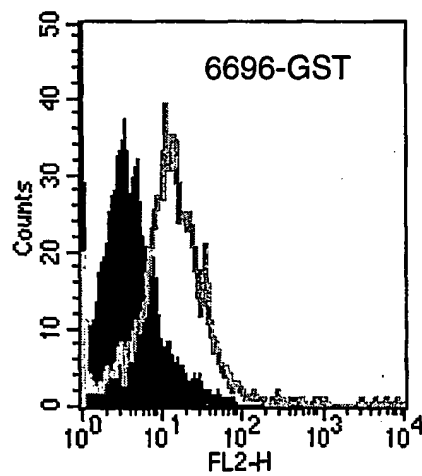
**Fig. 51A**



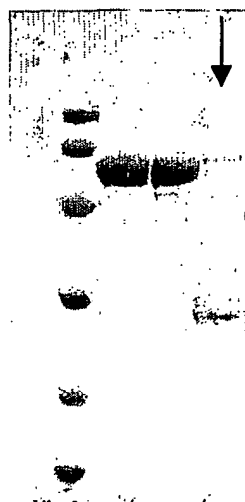
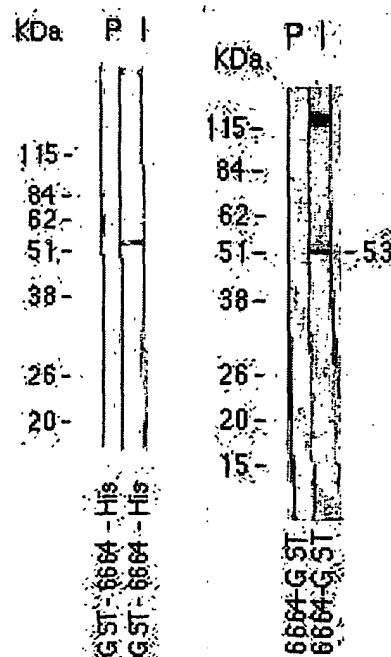
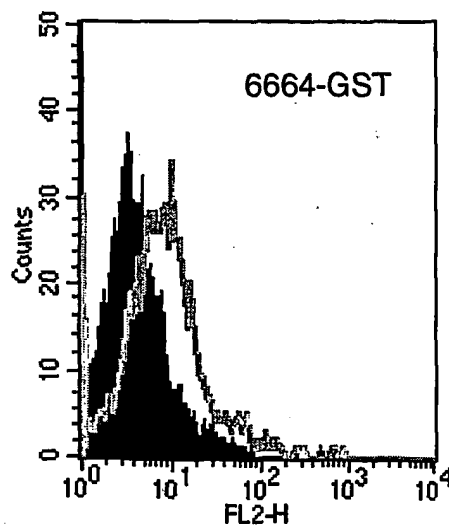
**Fig. 51B**



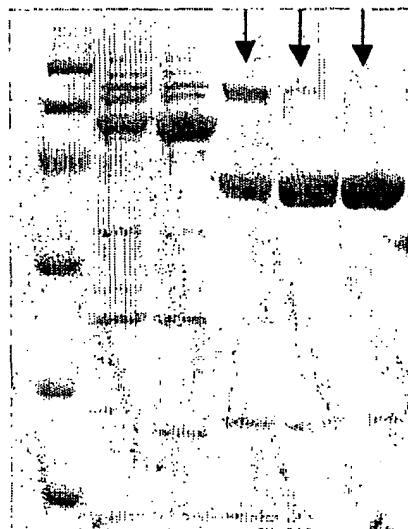
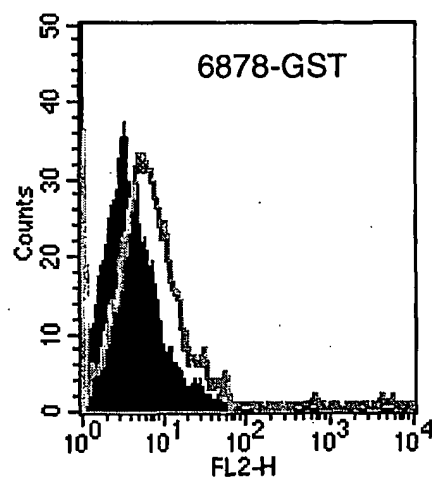
**Fig. 51C**



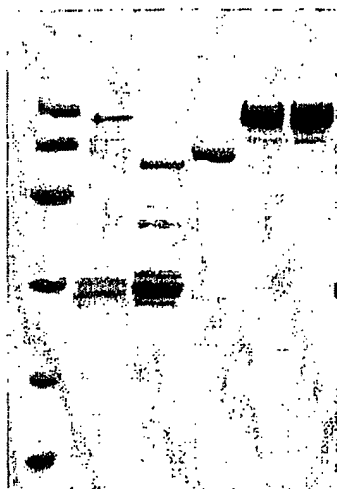
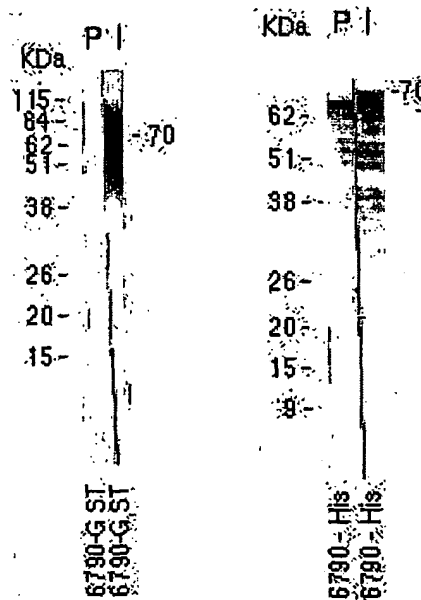
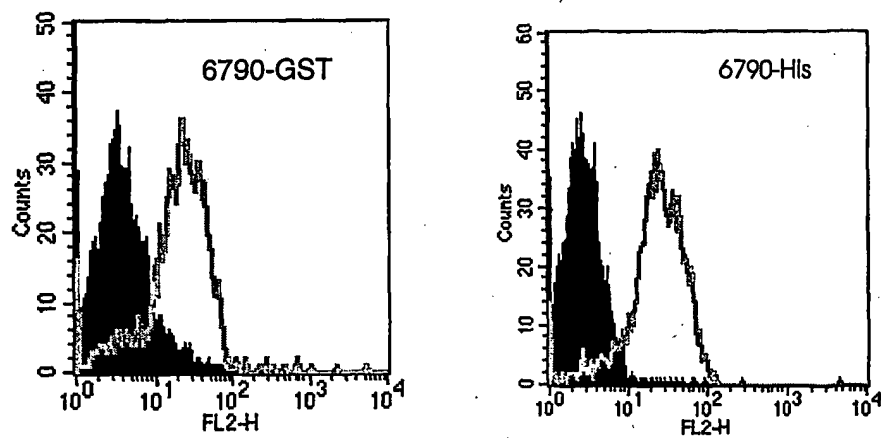
50/169

**FIGURE 50****Fig. 50A****Fig. 50B****Fig. 50C**

53/169

**FIGURE 53****Fig. 53A****Fig. 53B**

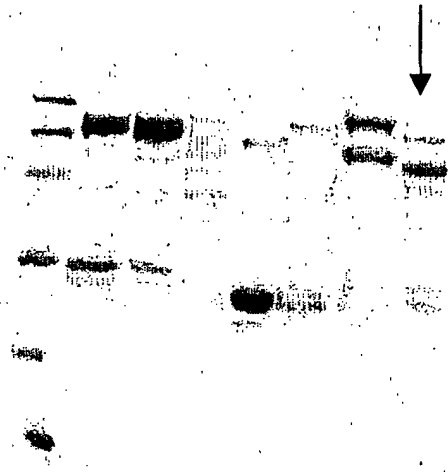
52/169

**FIGURE 52****Fig. 52A****Fig. 52B****Fig. 52C**

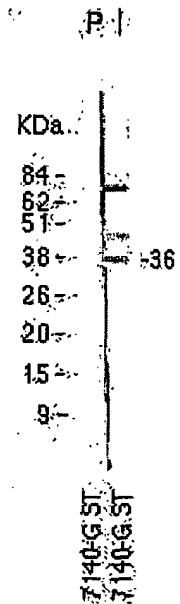
55/169

**FIGURE 55**

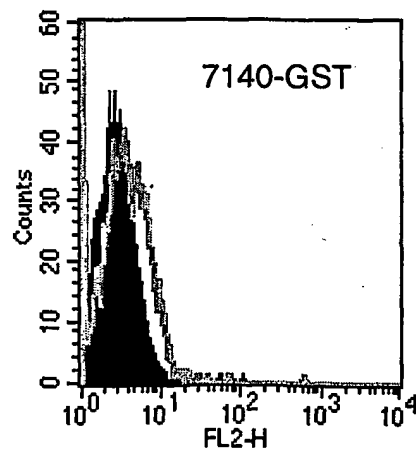
**Fig. 55A**



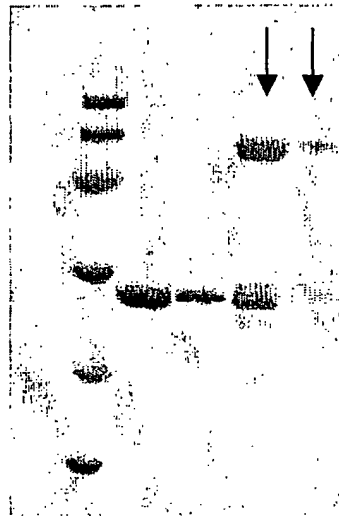
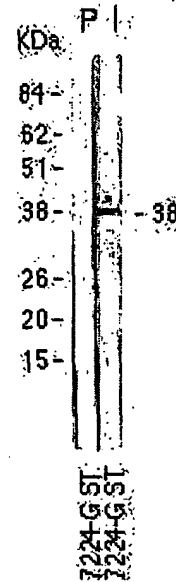
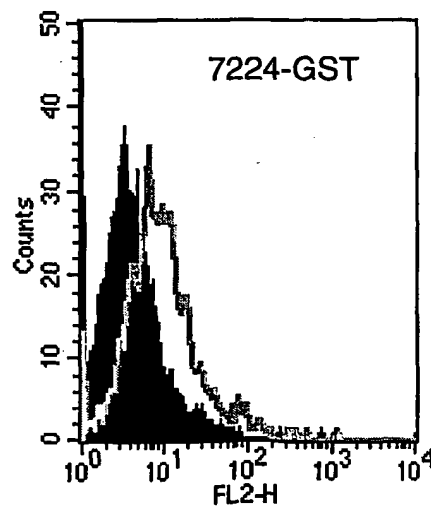
**Fig. 55B**



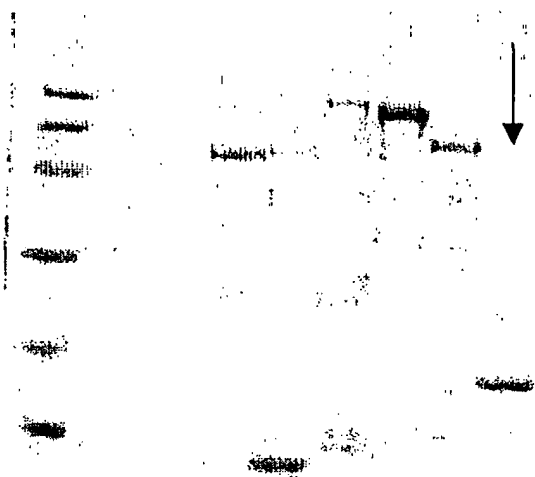
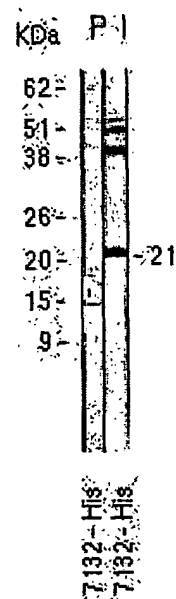
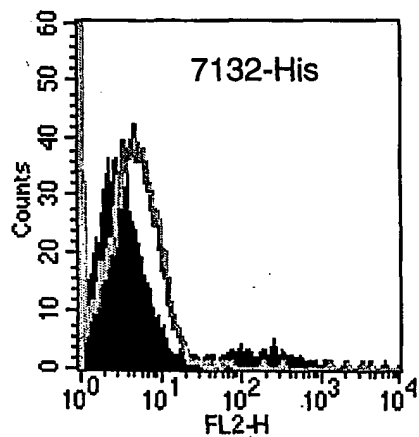
**Fig. 55C**



54/169

**FIGURE 54****FIG. 54A****FIG. 54B****FIG. 54C**

57/169

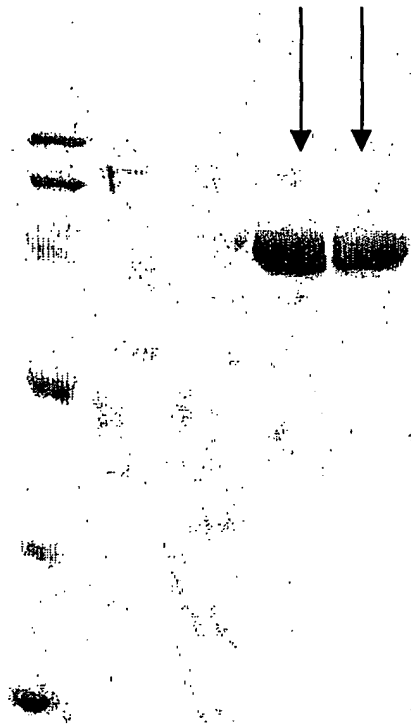
**FIGURE 57****Fig. 57A****Fig. 57B****Fig. 57C**



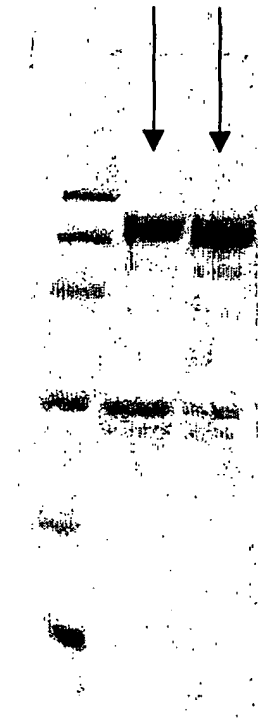
56/169

**FIGURE 56**

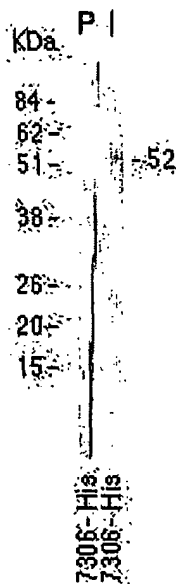
**Fig. 56A**



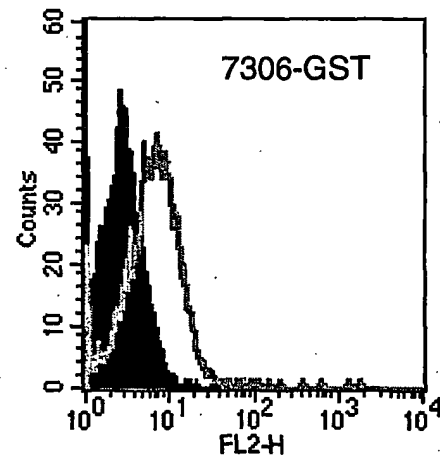
**Fig. 56B**



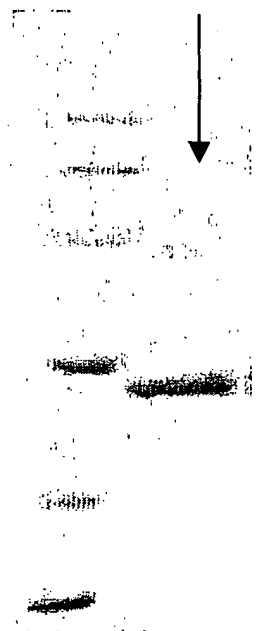
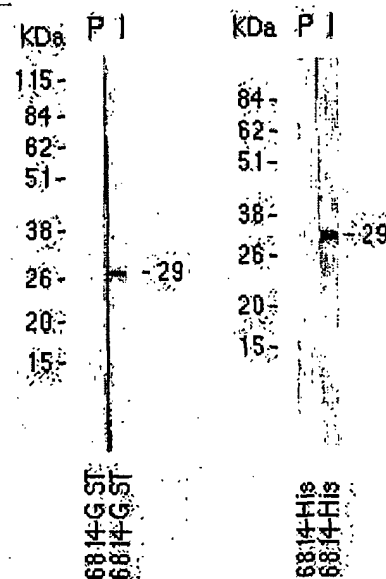
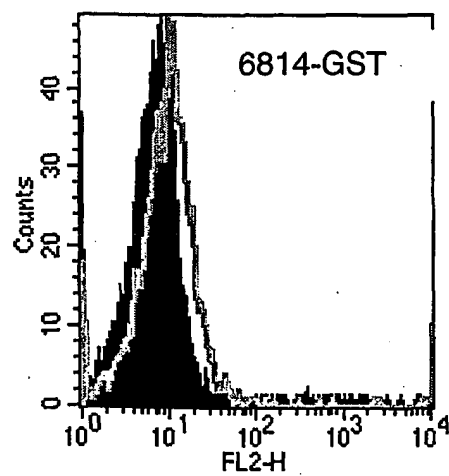
**FIG. 56C**



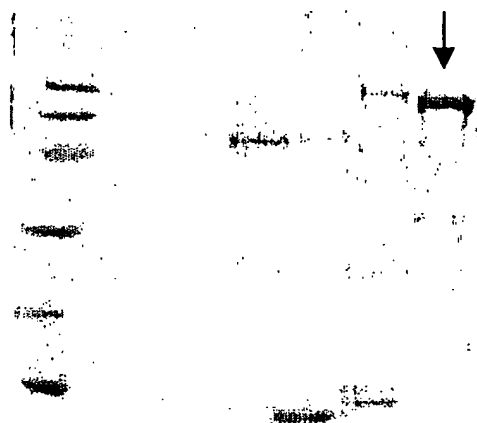
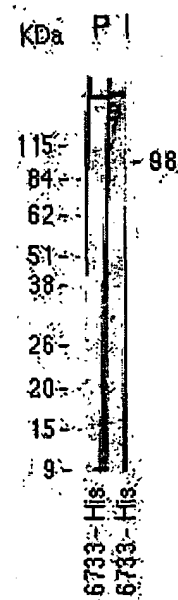
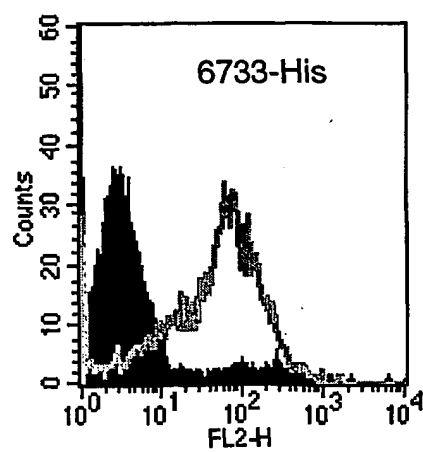
**FIG. 56D**



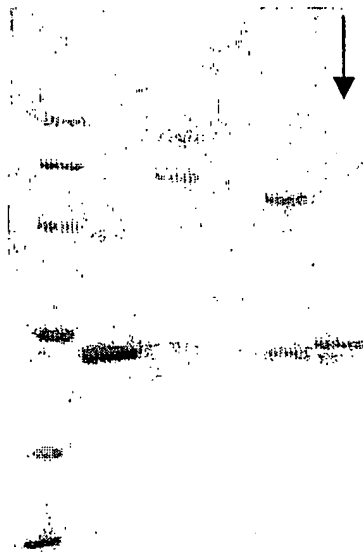
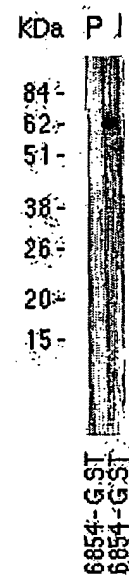
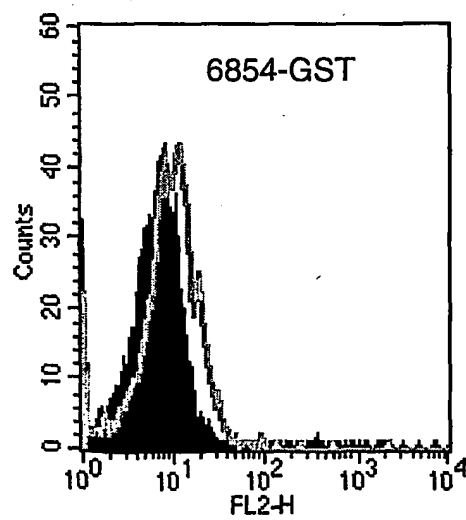
59/169

**FIGURE 59****Fig. 59A****Fig. 59B****Fig. 59C**

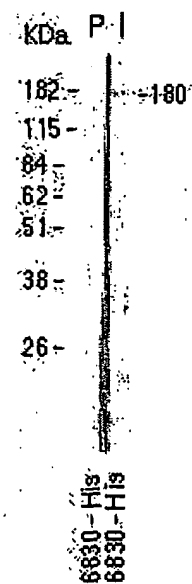
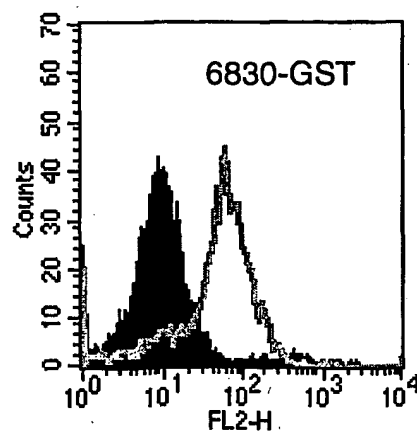
58/169

**FIGURE 58****FIG. 58A****Fig. 58B****FIG. 58C**

61/169

**FIGURE 61****Fig. 61A****Fig. 61B****Fig. 61C**

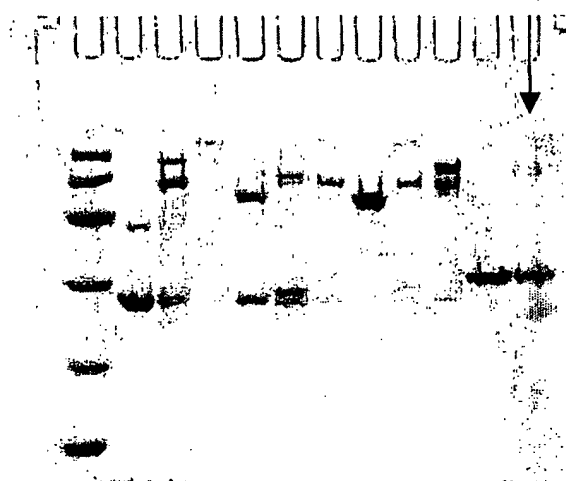
60/169

**FIGURE 60****FIG. 60A****FIG. 60B****FIG. 60C**

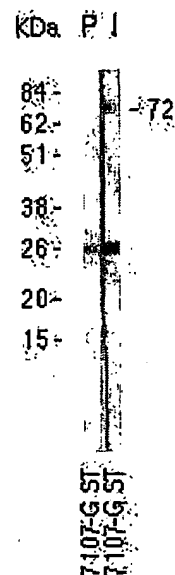
63/169

**FIGURE 63**

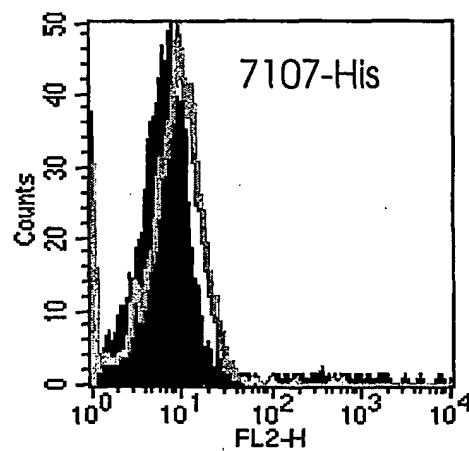
**Fig. 63A**



**Fig. 63B**



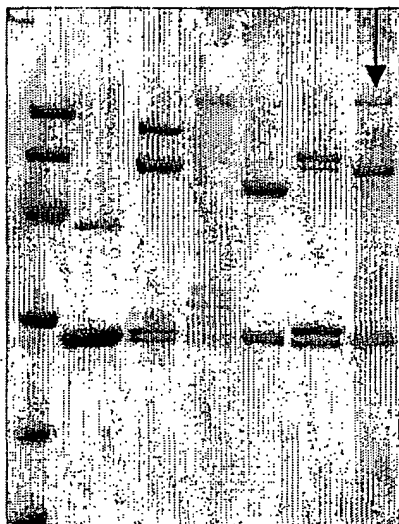
**Fig. 63C**



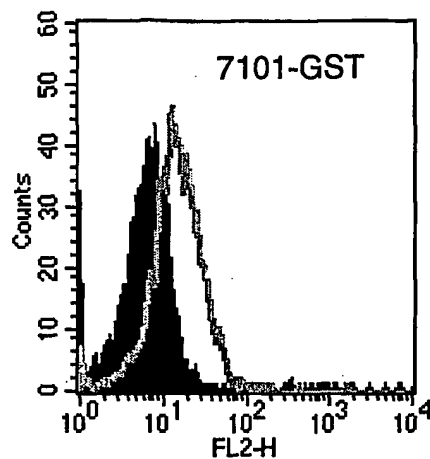
62/169

# **FIGURE 62**

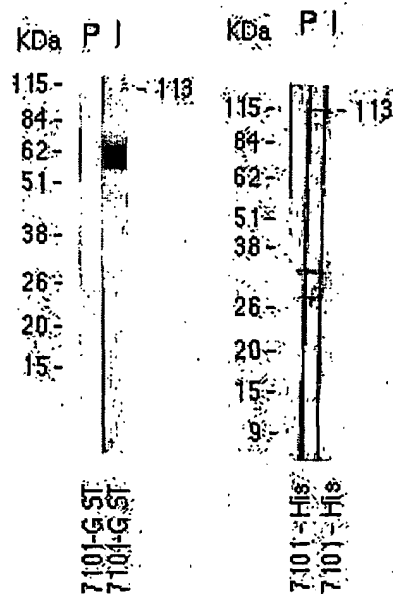
**Fig. 62A**



**Fig. 62C**



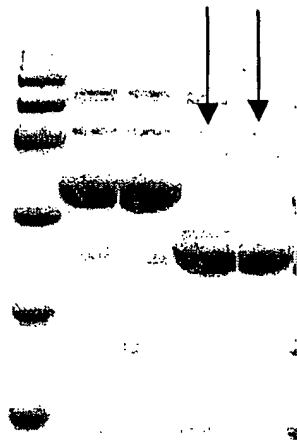
**Fig. 62B**



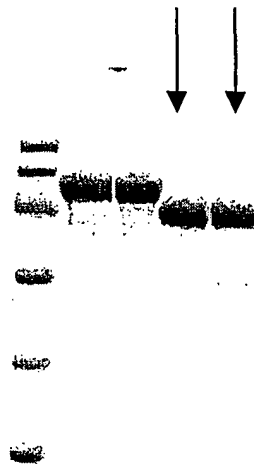
65/169

**FIGURE 65**

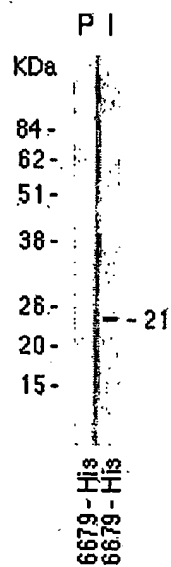
**Fig. 65A**



**Fig. 65B**



**Fig. 65C**





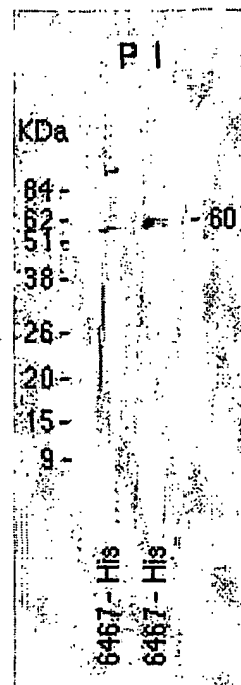
64/169

**FIGURE 64**

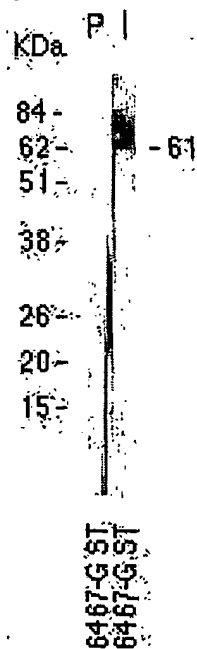
**FIG. 64A**



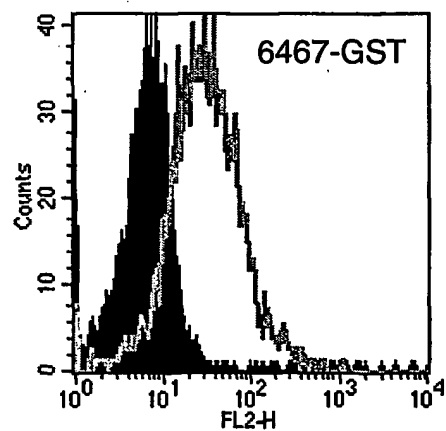
**FIG. 64B**



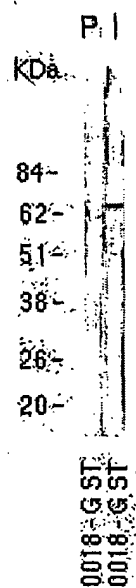
**FIG. 64C**



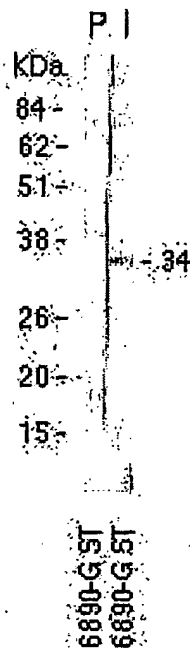
**FIG. 64D**



67/169

**FIGURE 67****Fig. 67A****Fig. 67B**

66/169

**FIGURE 66****FIG. 66A****FIG. 66B**

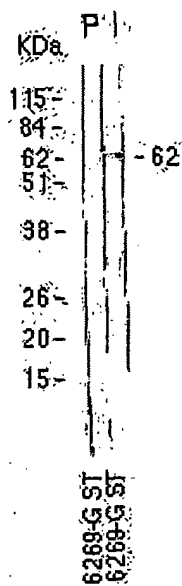
69/169

**FIGURE 69**

**Fig. 69A**



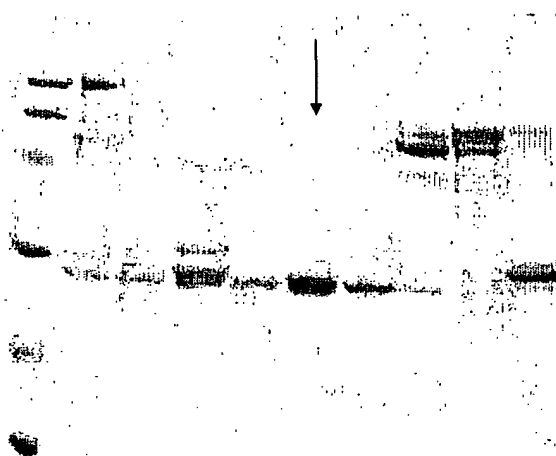
**Fig. 69B**



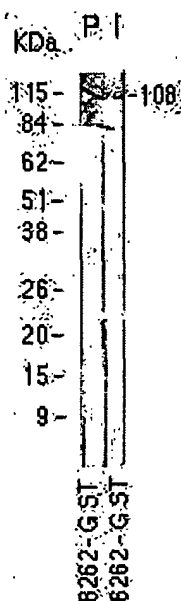
68/169

**FIGURE 68**

**FIG. 68A**



**FIG. 68B**



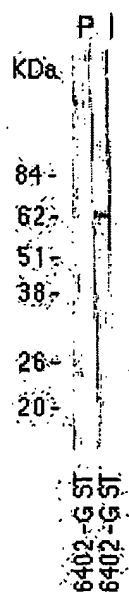
71/169

**FIGURE 71**

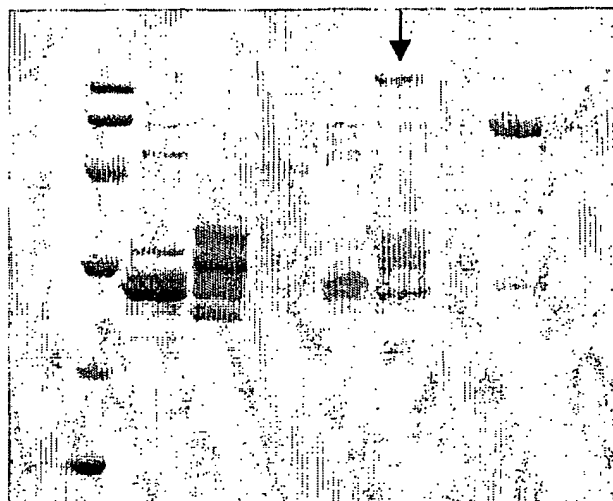
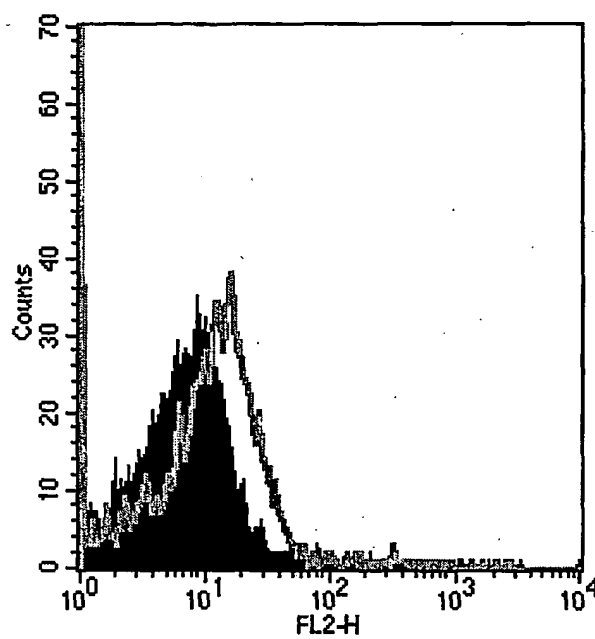
**FIG. 71A**



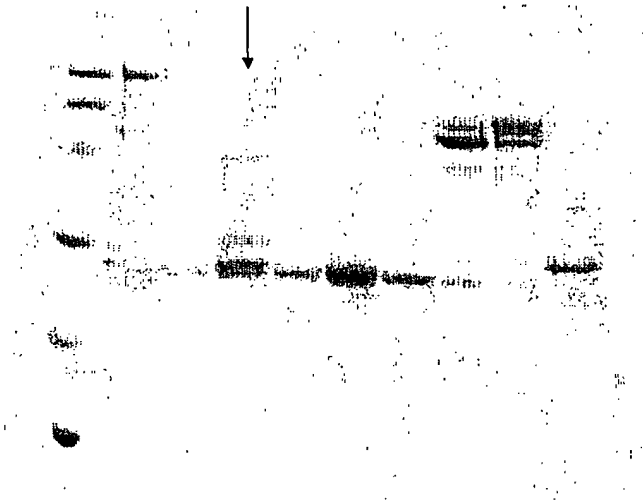
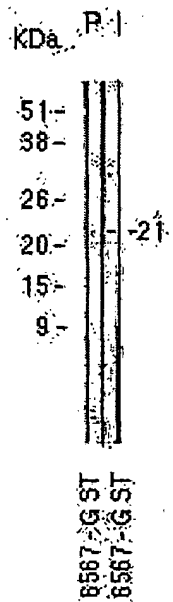
**FIG. 71B**



70/169

**FIGURE 70****FIG. 70A****FIG. 70B**

73/169

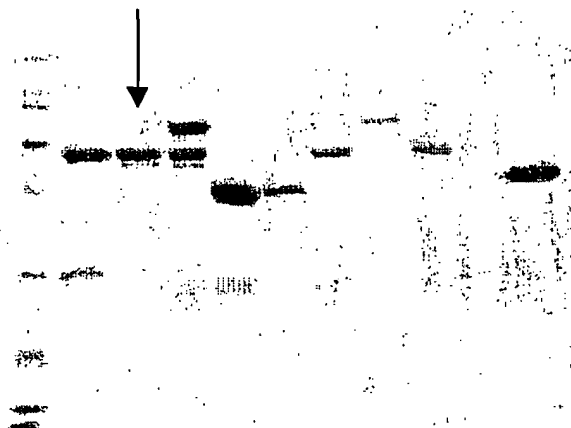
**FIGURE 73****FIG. 73A****FIG. 73B**



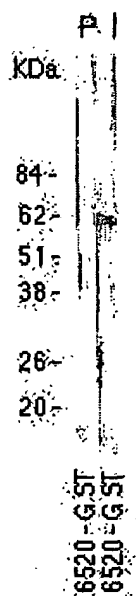
72/169

**FIGURE 72**

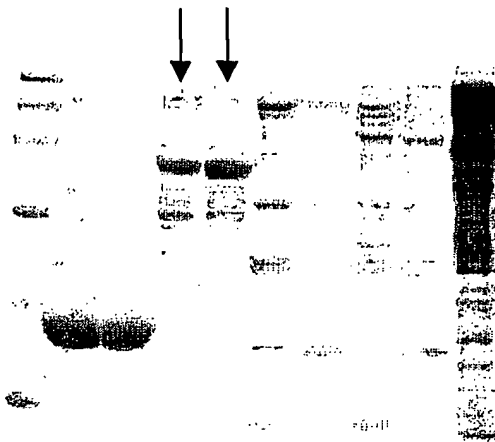
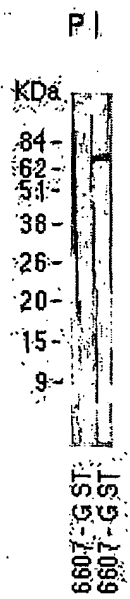
**FIG. 72A**



**FIG. 72B**



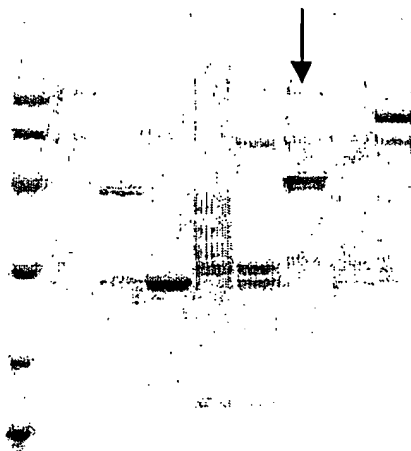
75/169

**FIGURE 75****Fig. 75A****Fig. 75B**

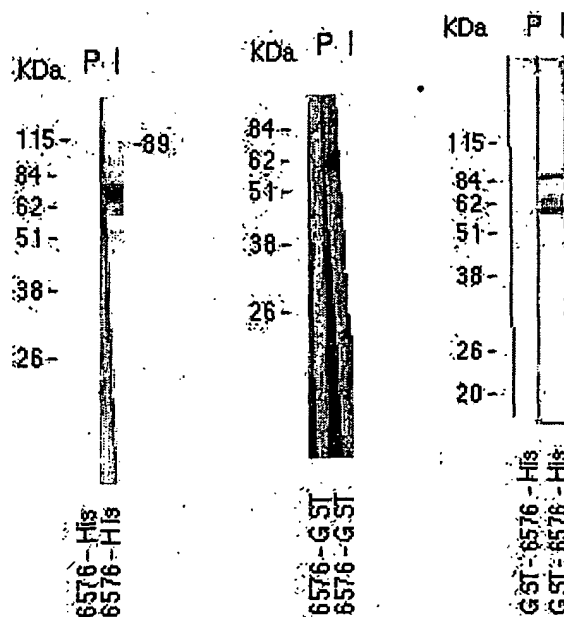
74/169

**FIGURE 74**

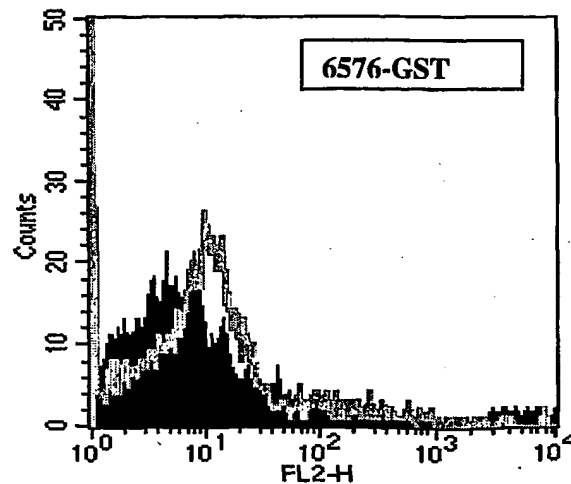
**Fig. 74A**



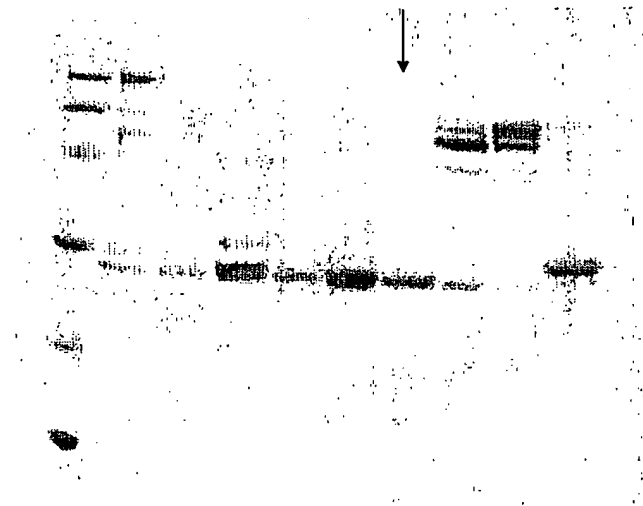
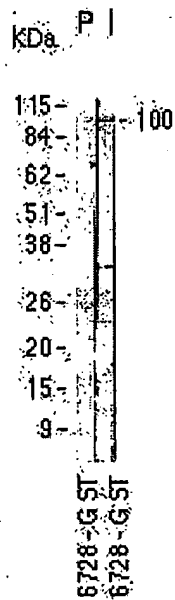
**Fig. 74B**



**Fig. 74C**



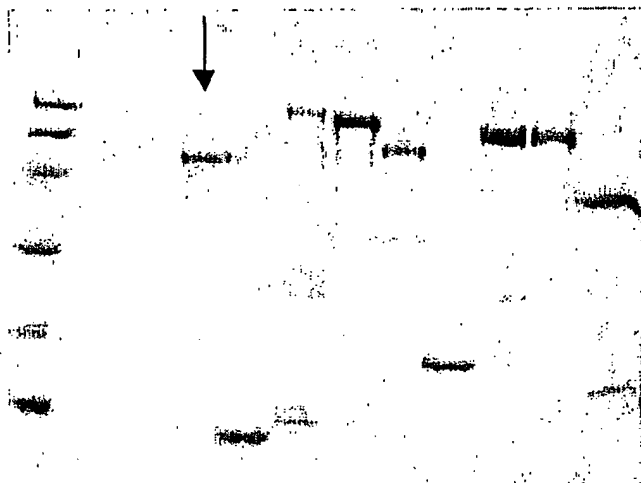
77/169

**FIGURE 77****FIG. 77A****FIG. 77B**

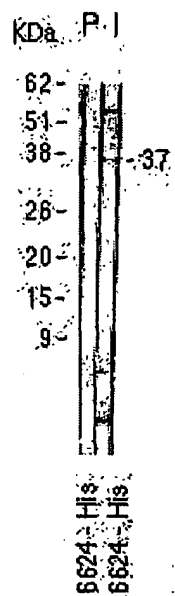
76/169

**FIGURE 76**

**Fig. 76A**



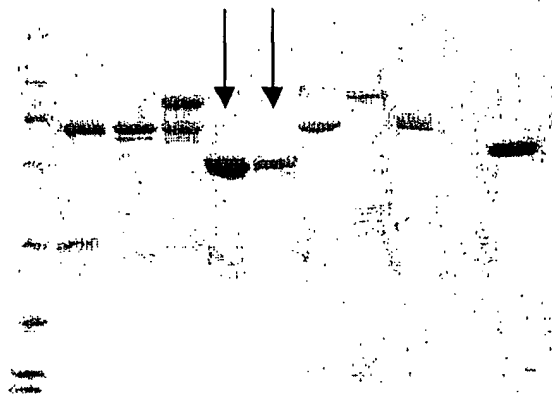
**Fig. 76B**



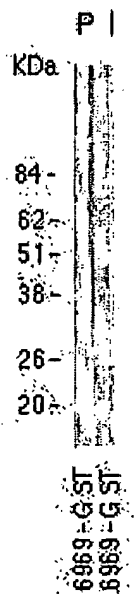
79/169

**FIGURE 79**

**FIG. 79A**



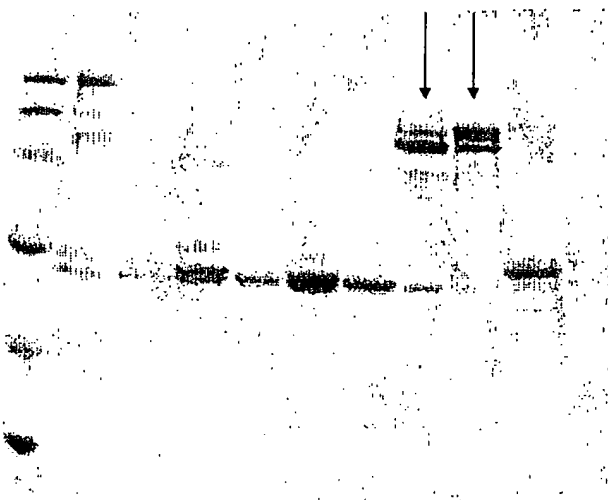
**FIG. 79B**



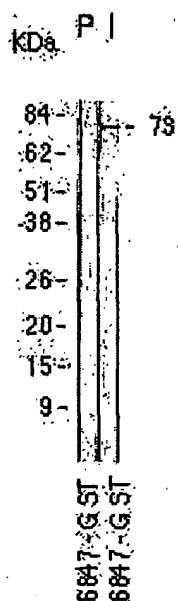
78/169

**FIGURE 78**

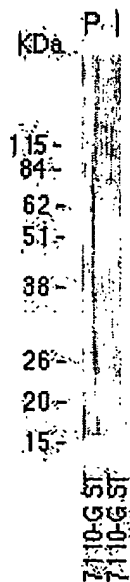
**FIG. 78A**



**FIG. 78B**



81/169

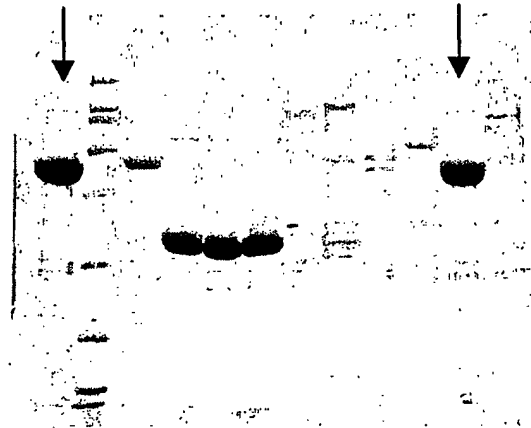
**FIGURE 81****FIG. 81A****FIG. 81B**



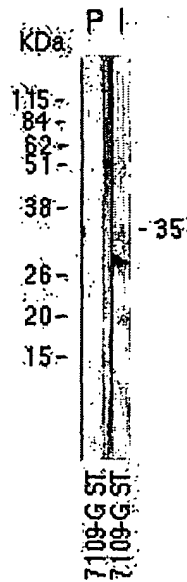
80/169

**FIGURE 80**

**FIG. 80A**



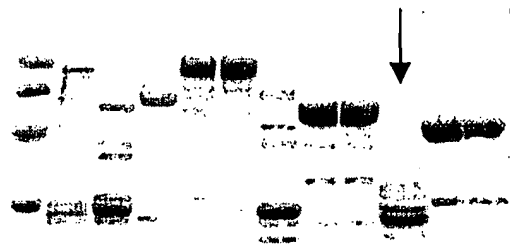
**FIG. 80B**



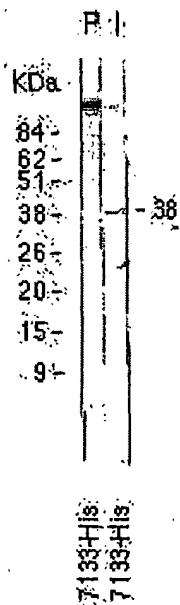
83/169

**FIGURE 83**

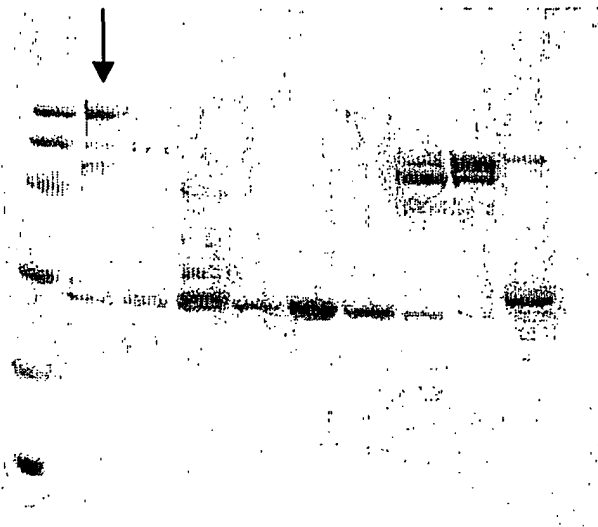
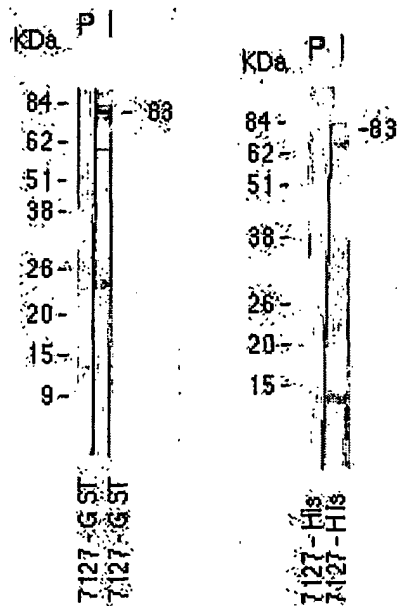
**Fig. 83A**



**Fig. 83B**



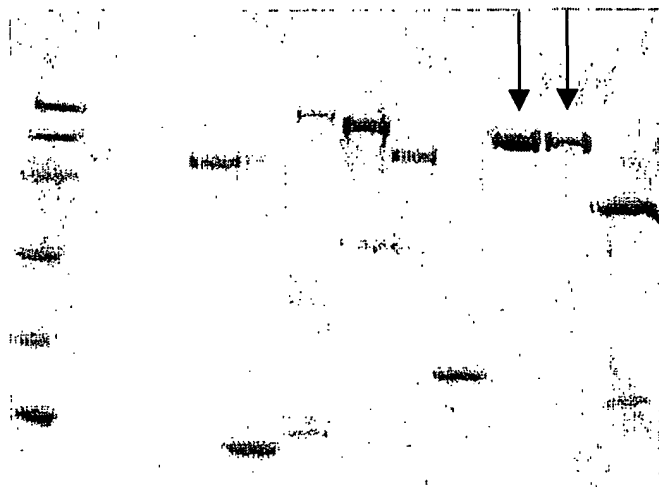
82/169

**FIGURE 82****Fig. 82A****Fig. 82B**

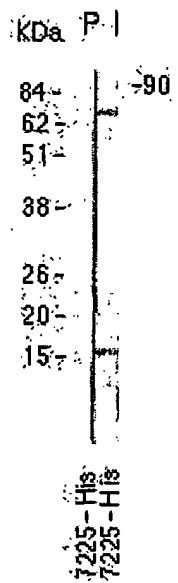
85/169

**FIGURE 85**

**Fig. 85A**



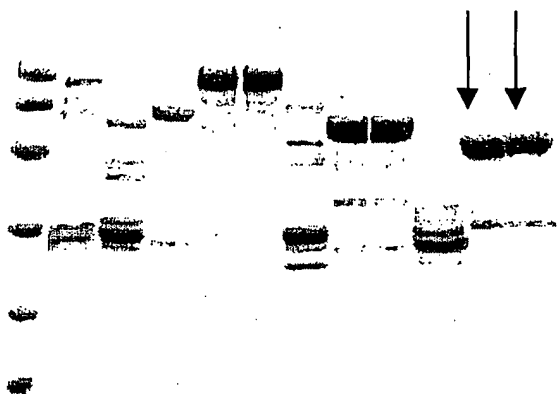
**Fig. 85B**



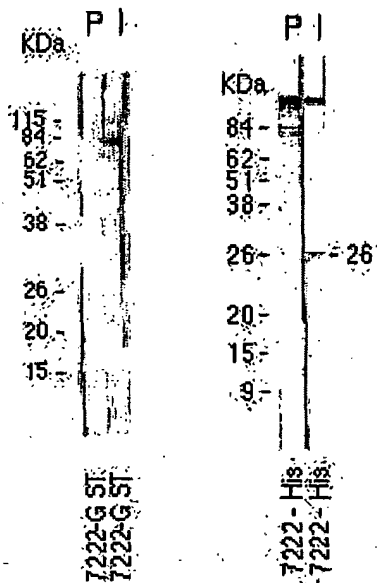
84/169

**FIGURE 84**

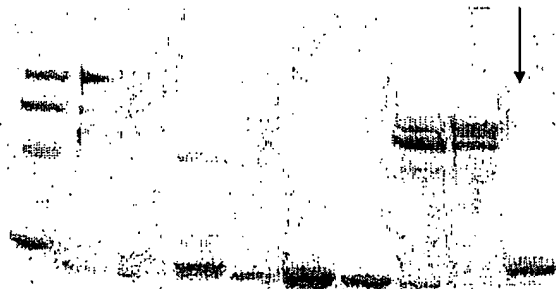
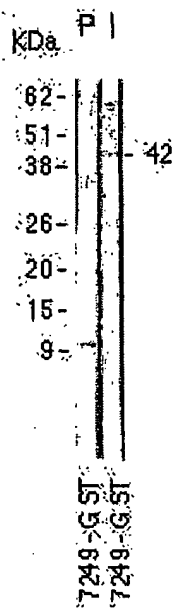
**FIG. 84A**



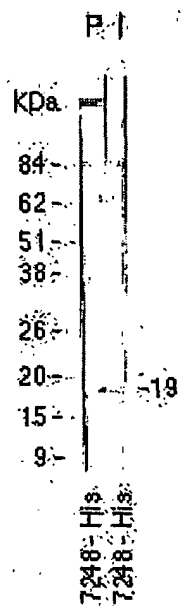
**FIG. 84B**



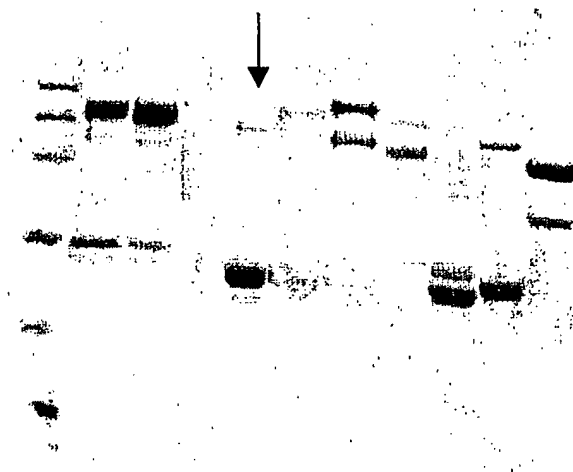
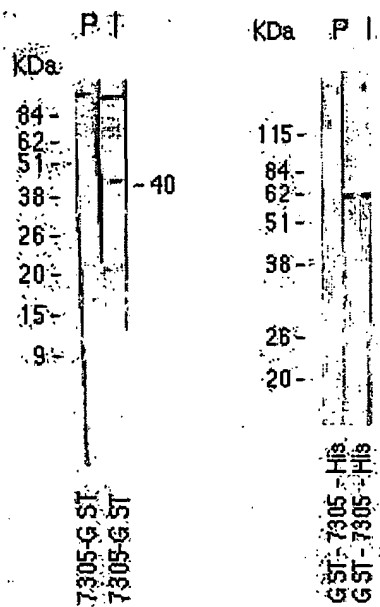
87/169

**FIGURE 87****Fig. 87A****Fig. 87B**

86/169

**FIGURE 86****FIG. 86A****FIG. 86B**

89/169

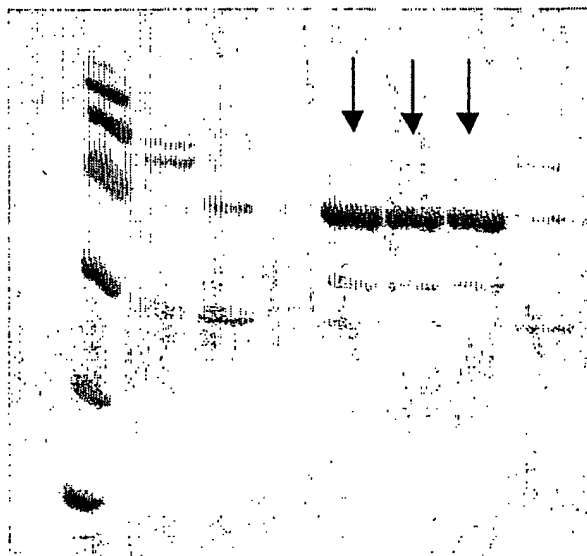
**FIGURE 89****FIG. 89A****FIG. 89B**



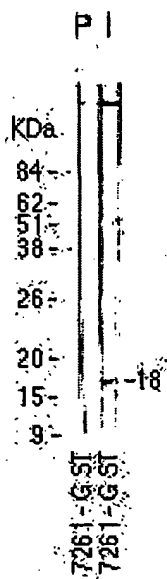
88/169

**FIGURE 88**

**FIG. 88A**



**FIG. 88B**



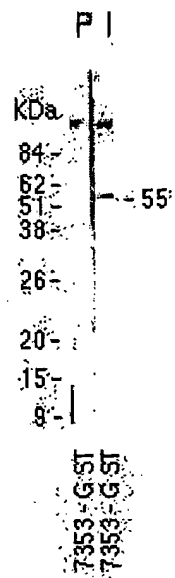
91/169

**FIGURE 91**

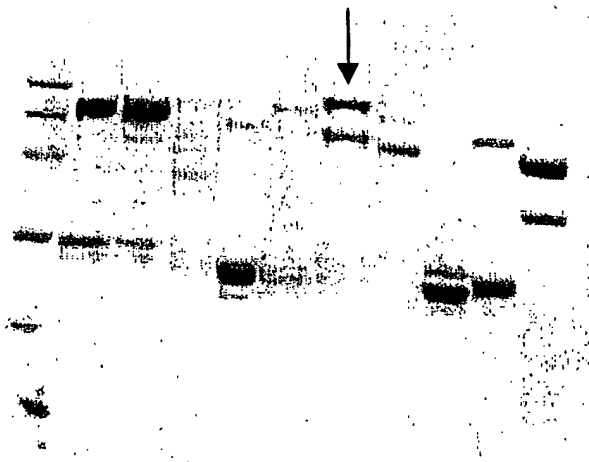
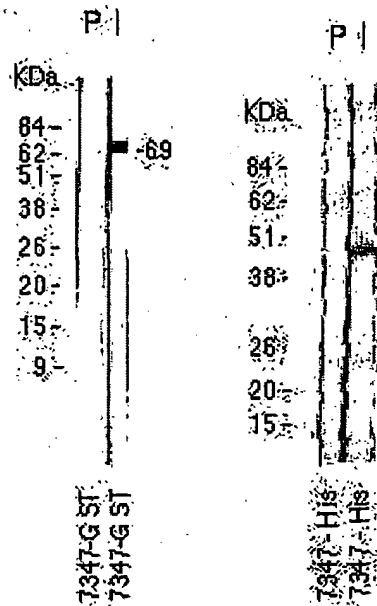
**Fig. 91A**



**Fig. 91B**



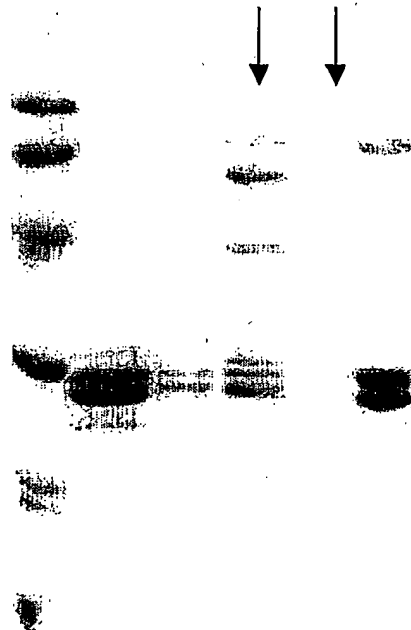
90/169

**FIGURE 90****Fig. 90A****Fig. 90B**

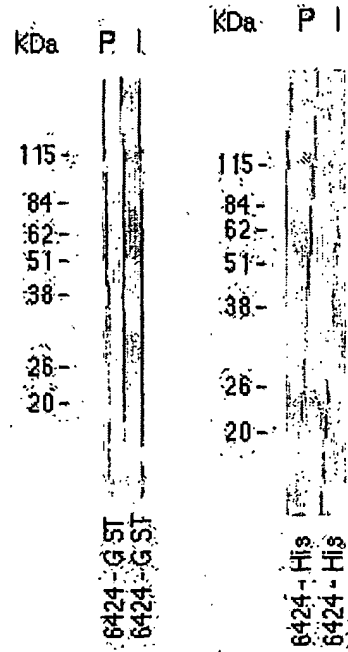
93/169

# **FIGURE 93**

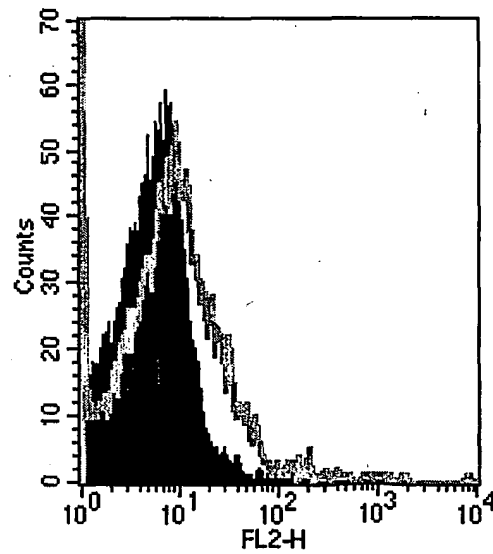
**FIG. 93A**



**FIG. 93B**



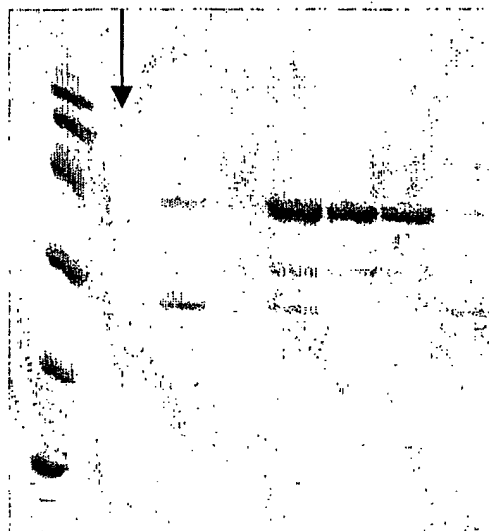
**FIG. 93C**



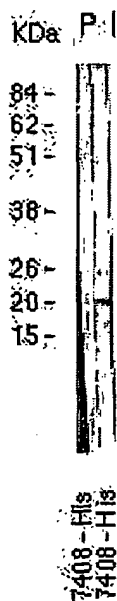
92/169

**FIGURE 92**

**FIG. 92A**



**FIG. 92B**



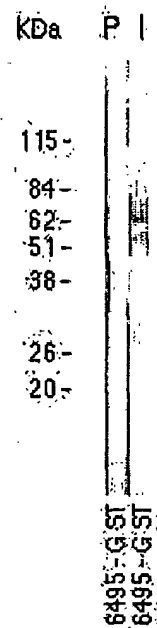
95/169

**FIGURE 95**

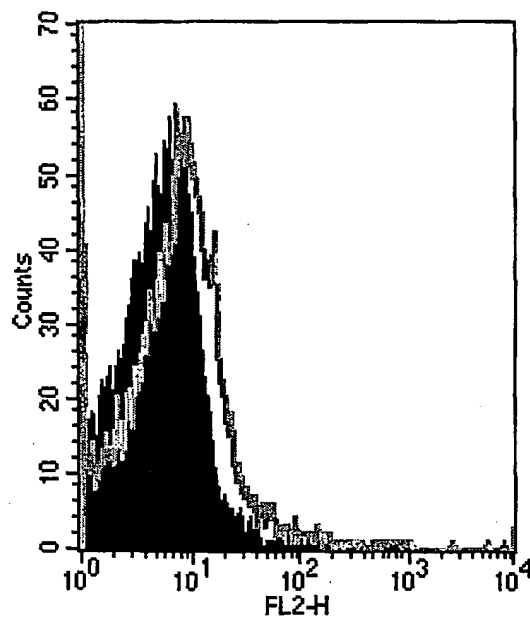
**Fig. 95A**



**Fig. 95B**



**FIG. 95C**



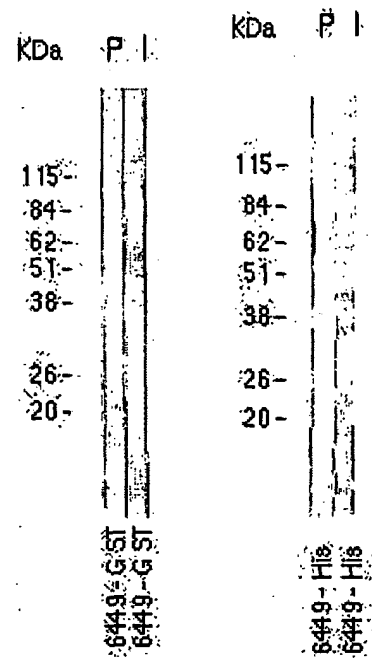
94/169

# **FIGURE 94**

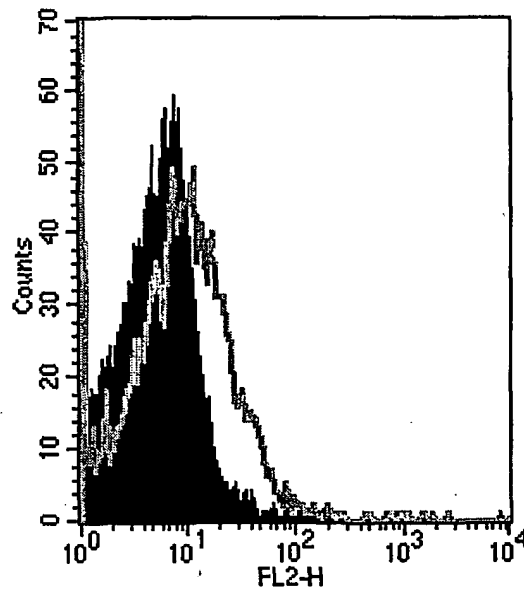
**FIG. 94A**



**FIG. 94B**



**FIG. 94C**



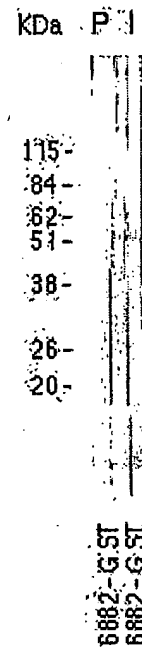
97/169

**FIGURE 97**

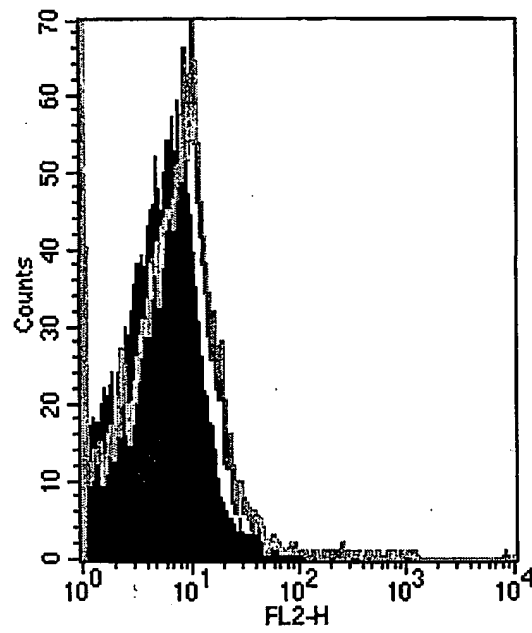
**Fig. 97A**



**Fig. 97B**

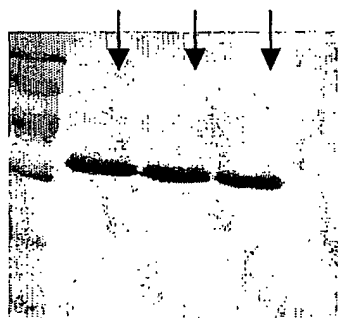
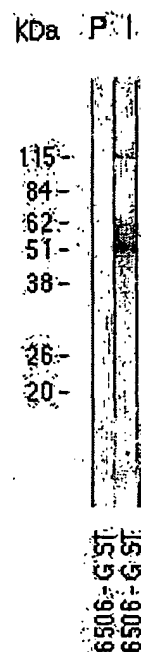
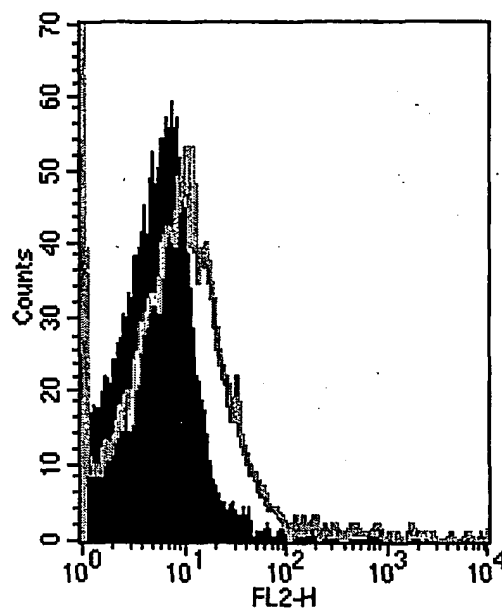


**Fig. 97C**

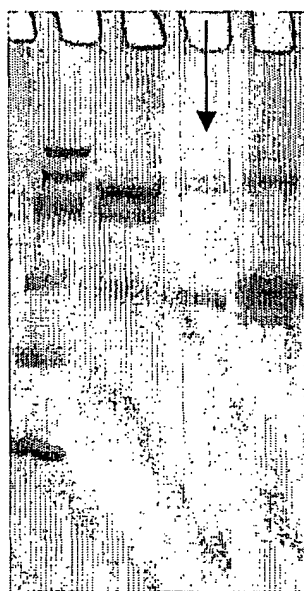
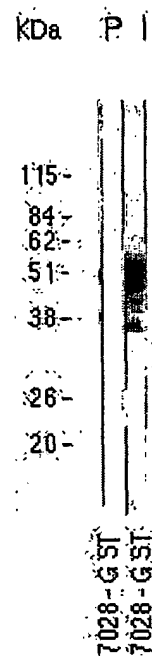
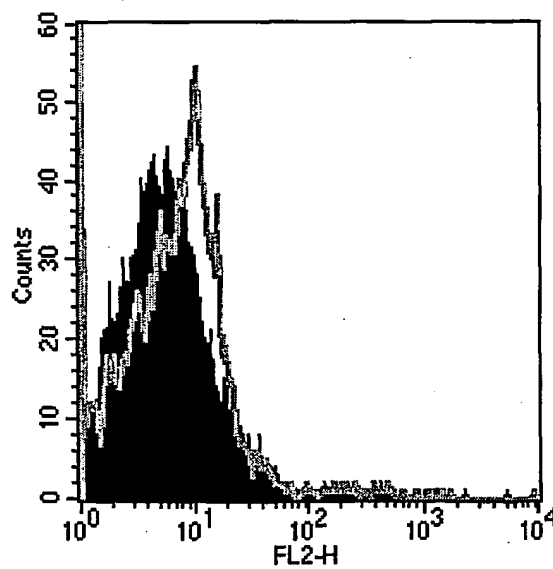




96/169

**FIGURE 96****FIG.  
96A****FIG.  
96B****FIG.  
96C****FIG. 96D**

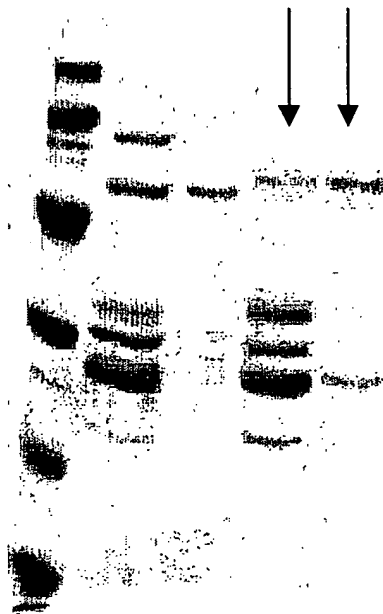
99/169

**FIGURE 99****FIG. 99A****FIG. 99B****FIG. 99C**

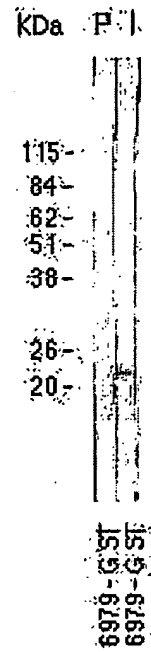
98/169

**FIGURE 98**

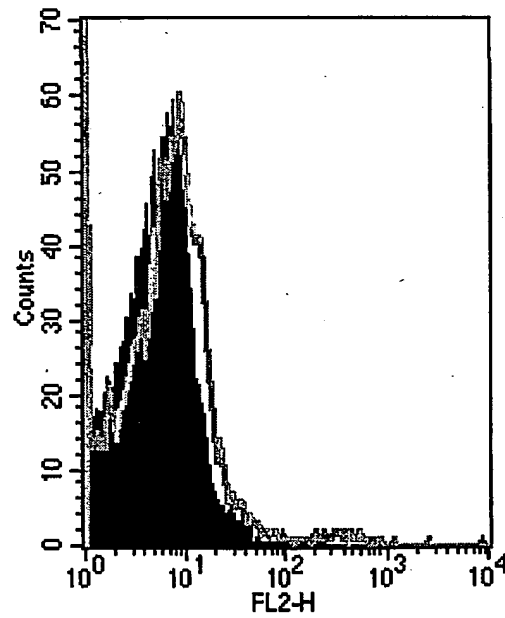
**FIG. 98A**



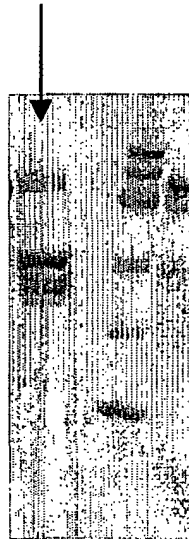
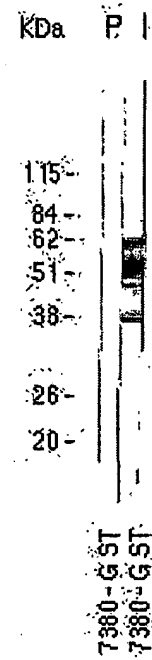
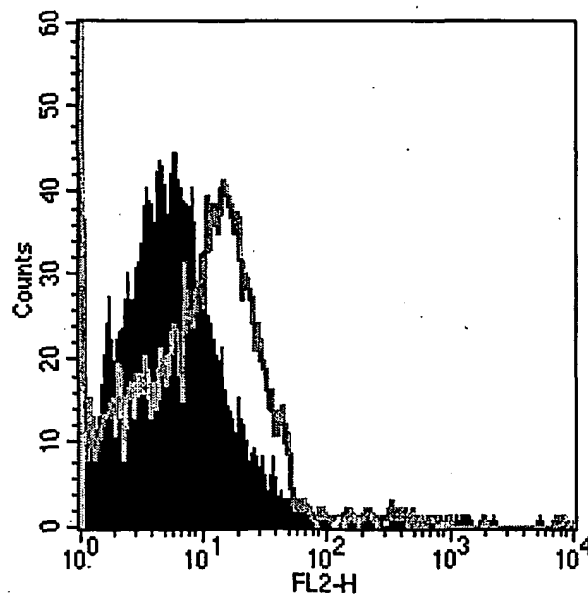
**FIG. 98B**



**FIG. 98C**



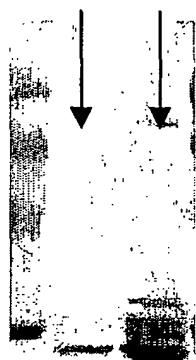
101/169

**FIGURE 101****FIG. 101A****FIG. 101B****FIG. 101C**

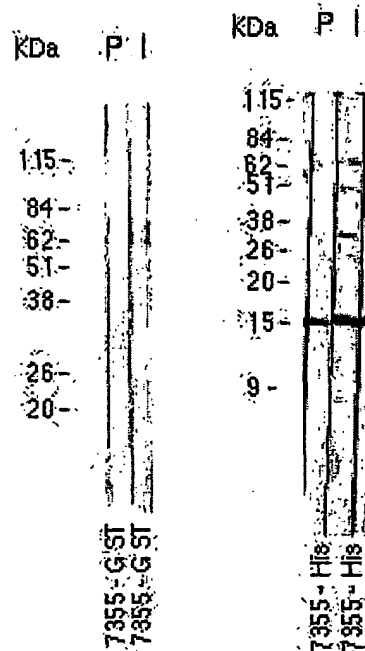
100/169

**FIGURE 100**

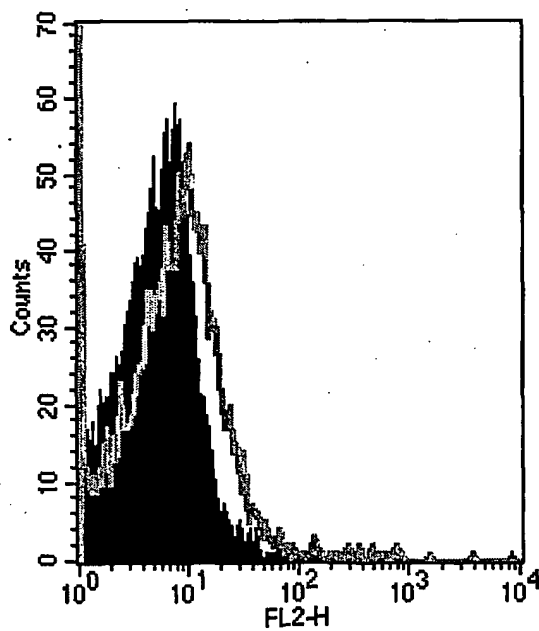
**Fig. 100A**



**Fig. 100B**

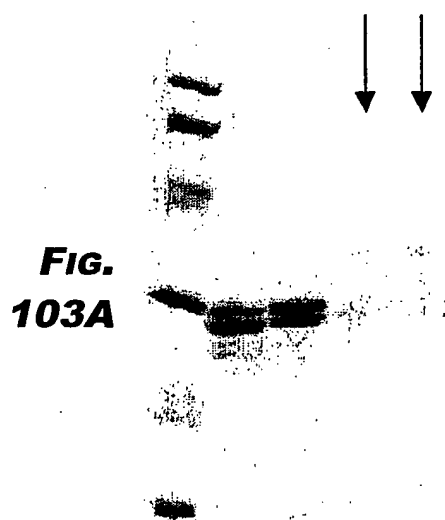


**Fig. 100C**

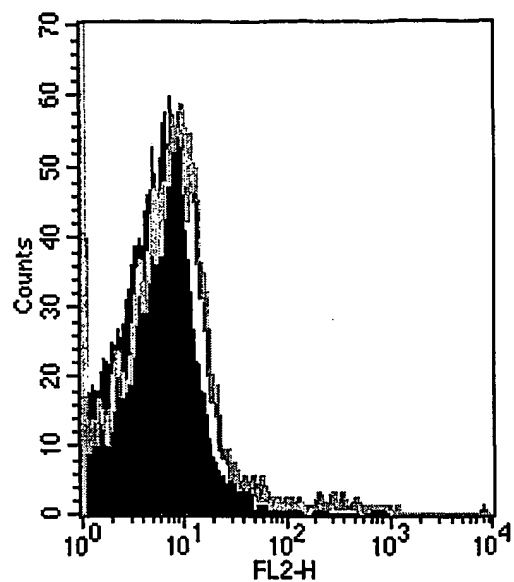


103/169

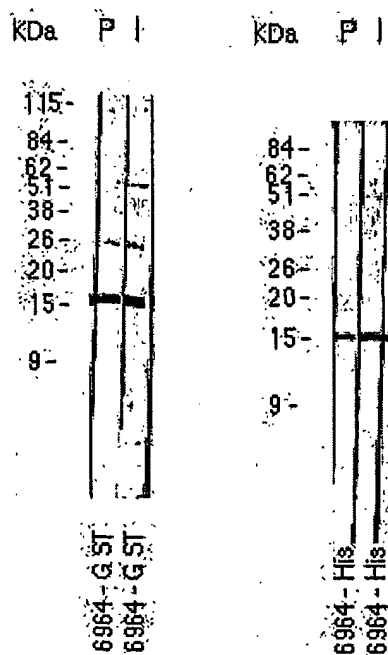
**FIGURE 103**



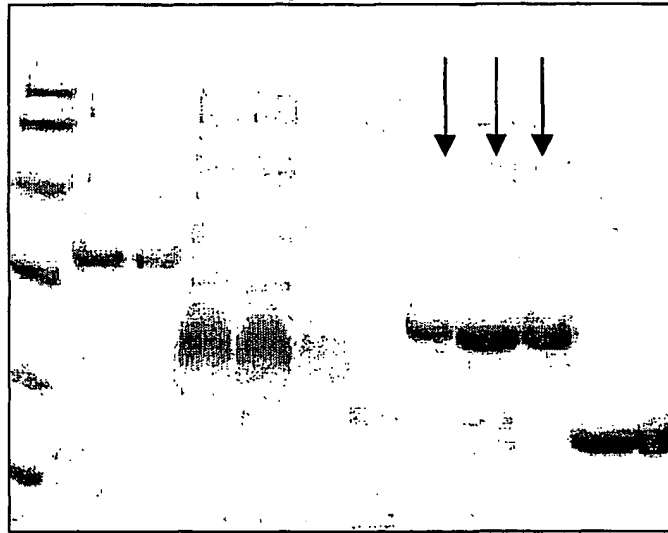
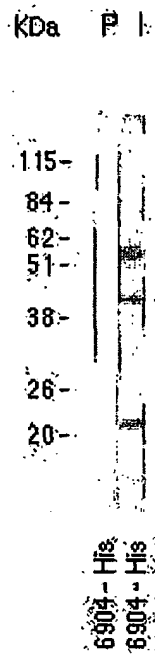
**FIG. 103C**



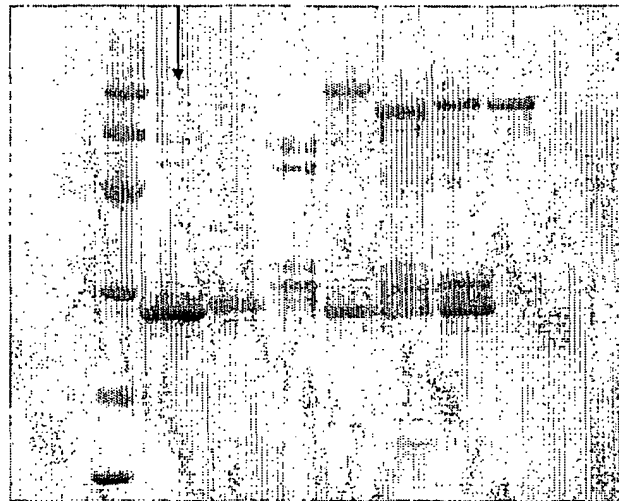
**FIG. 103B**



102/169

**FIGURE 102****FIG. 102A****FIG. 102B**

105/169

**FIGURE 105****FIG. 105A**

kDa P I

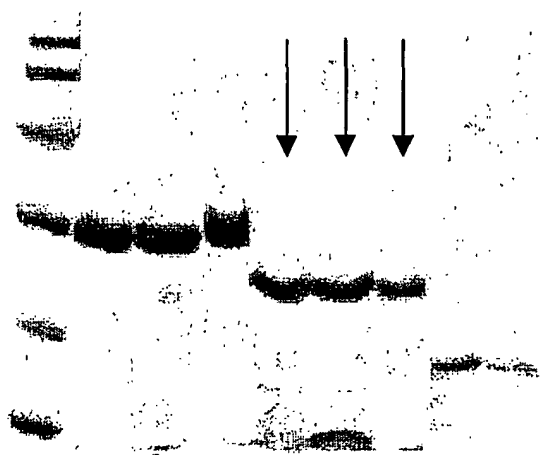
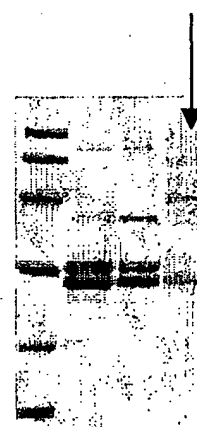
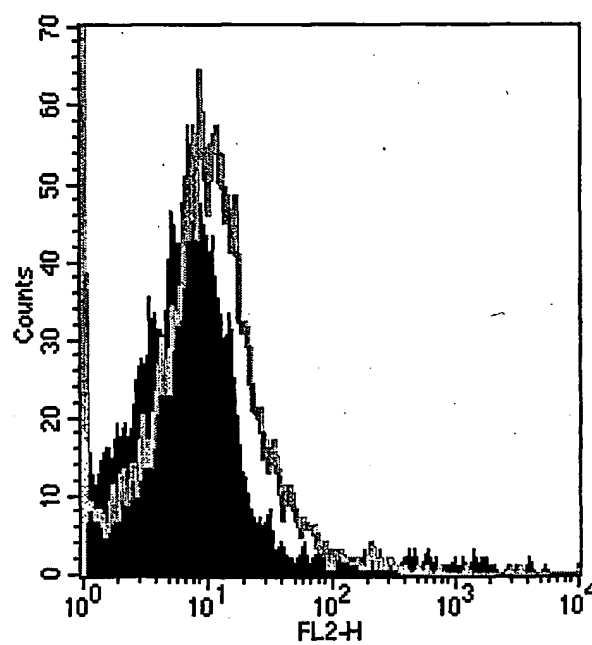
**FIG. 105B**

115-  
84-  
62-  
51-  
38-  
26-  
20-

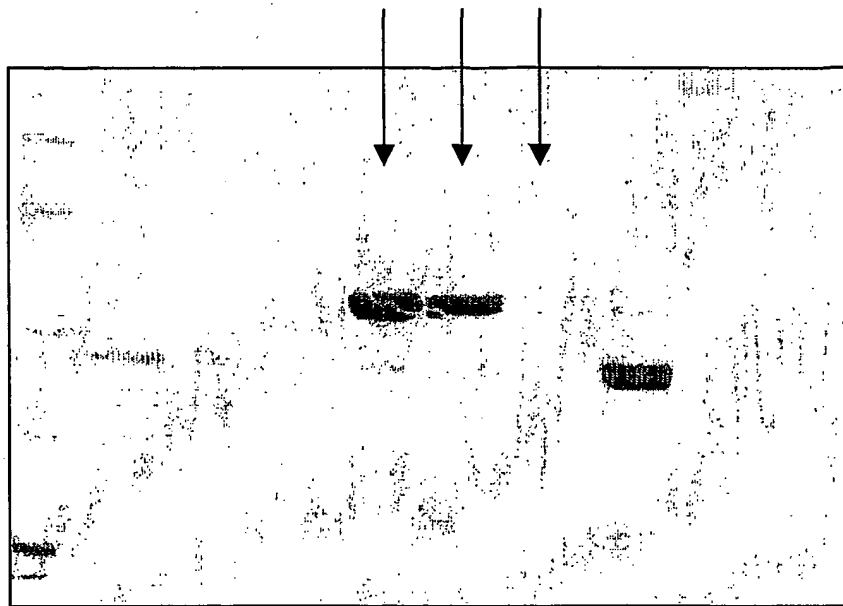
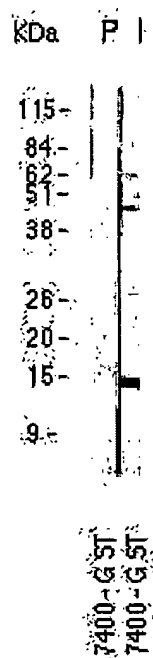
6281-G ST  
6281-G ST



104/169

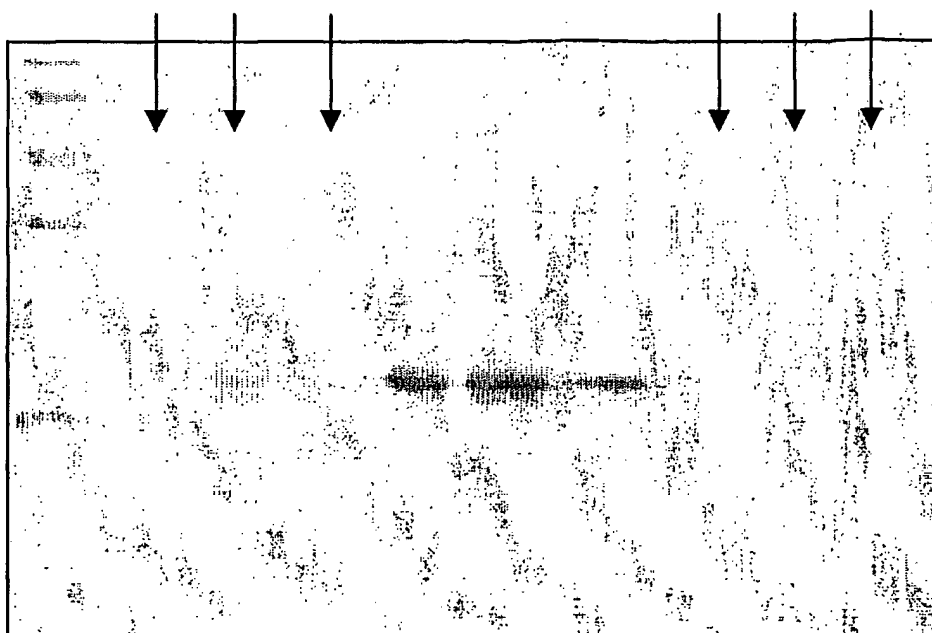
**FIGURE 104****FIG. 104A****FIG. 104B****FIG. 104C**

107/169

**FIGURE 108****Fig. 108A****Fig. 108B**

106/169

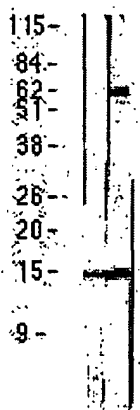
**FIGURE 106**



**Fig. 106A**

**Fig. 106B**

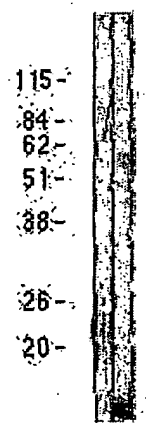
KDa P I.



His  
6306  
His  
6306

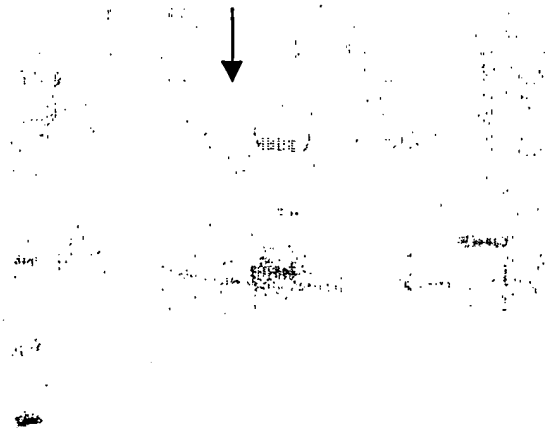
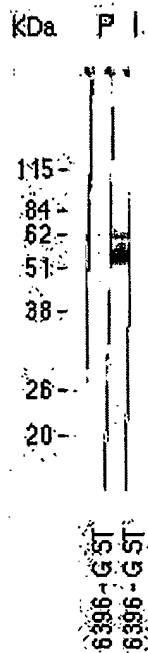
**FIGURE 107**

KDa P I.



His  
6434  
His  
6434

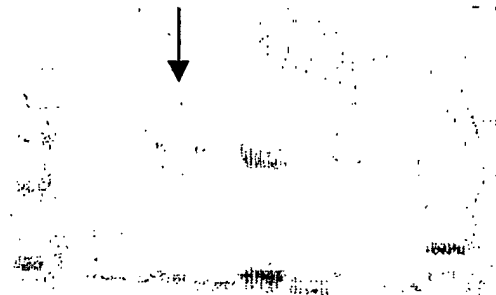
109/169

**FIGURE 110****Fig. 110A****Fig. 110B**

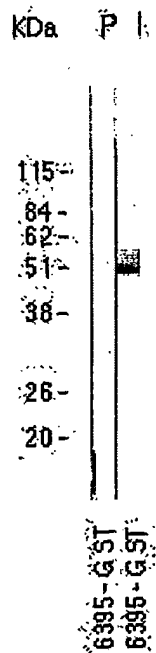
108/169

**FIGURE 109**

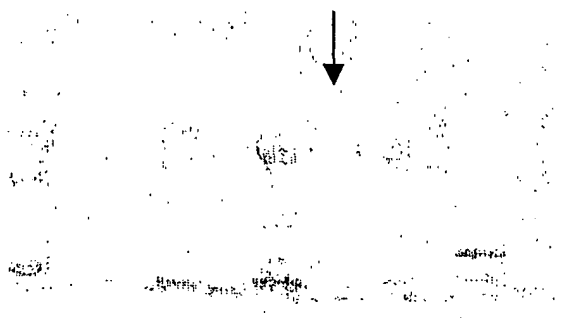
**Fig. 109A**



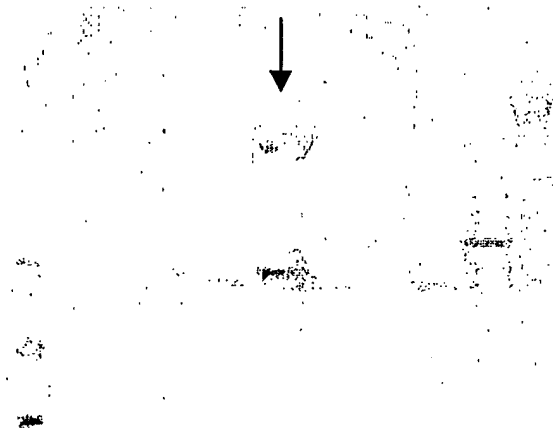
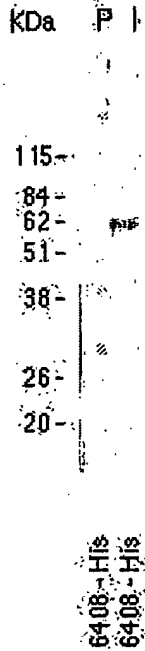
**Fig. 109B**



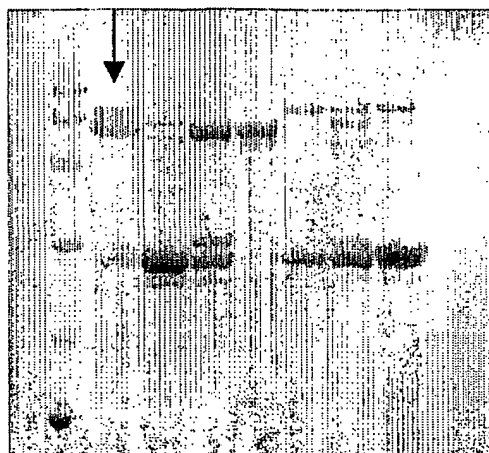
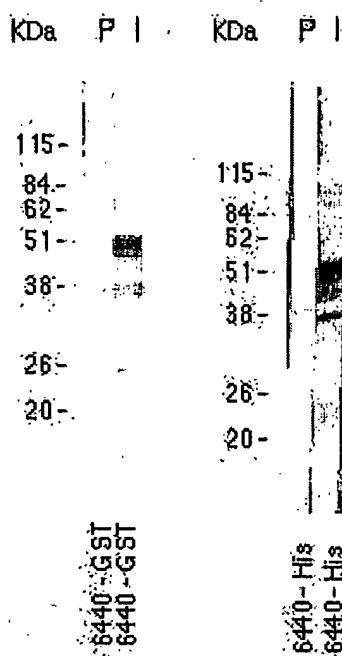
111/169

**FIGURE 112****FIG. 112A****FIG. 112B**

110/169

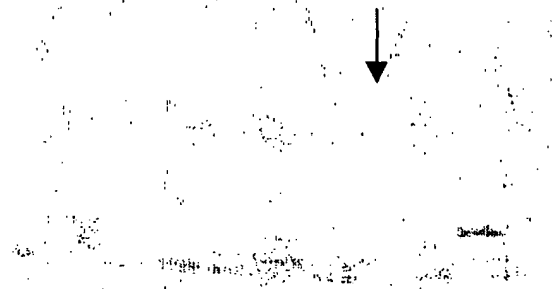
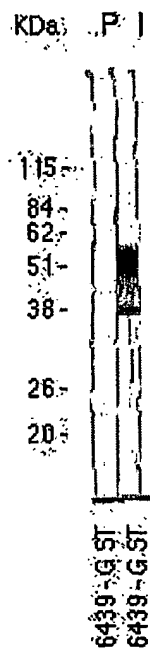
**FIGURE 111****FIG. 111A****FIG. 111B**

113/169

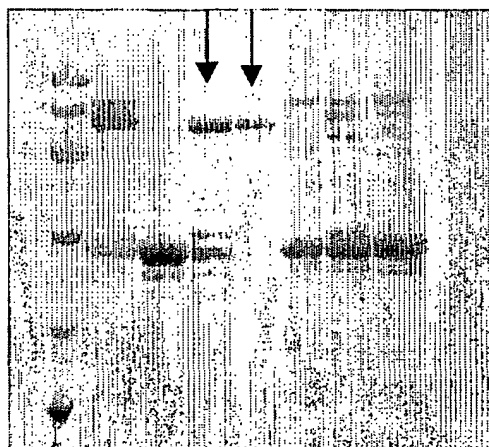
**FIGURE 114****FIG. 114A****FIG. 114B**



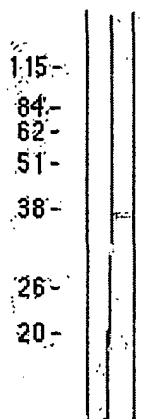
112/169

**FIGURE 113****Fig. 113A****Fig. 113B**

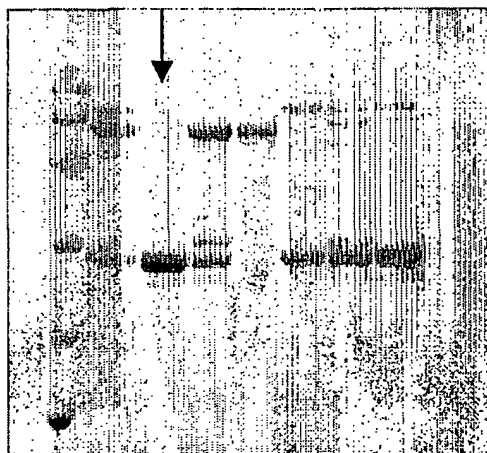
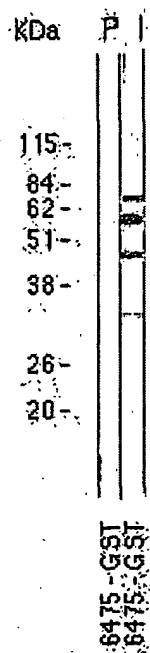
115/169

**FIGURE 116****Fig. 116A****Fig. 116B**

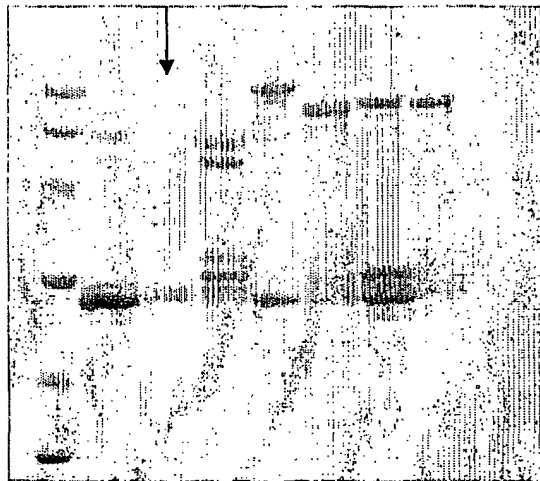
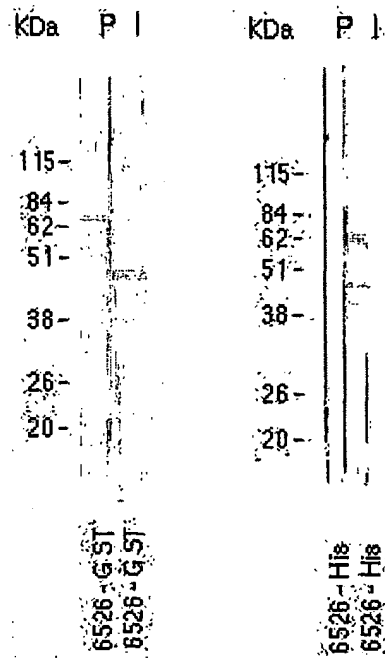
kDa P |

6482 - GST  
6482 - GST

114/169

**FIGURE 115****FIG. 115A****FIG. 115B**

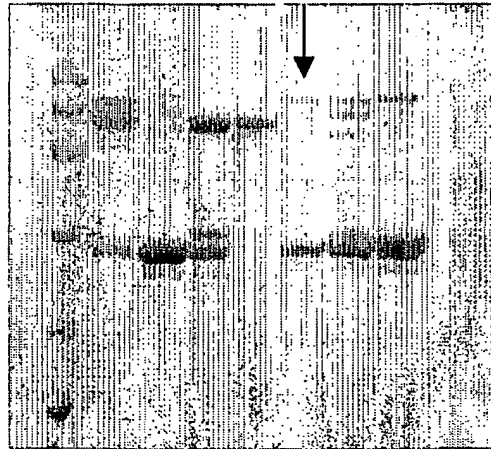
117/169

**FIGURE 118****Fig. 118A****Fig. 118B**

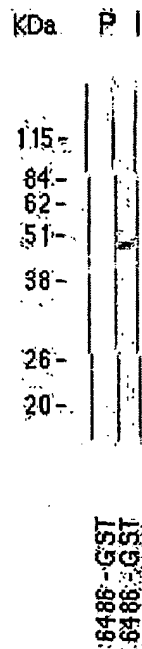
116/169

**FIGURE 117**

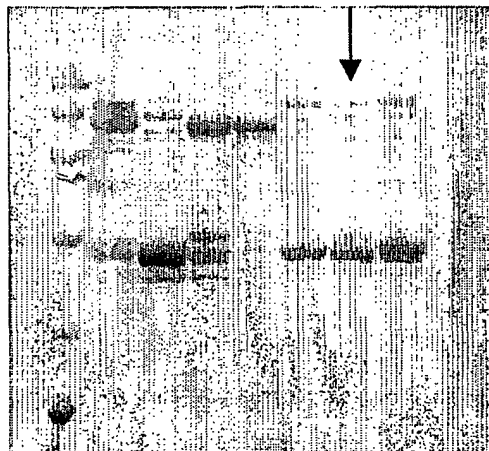
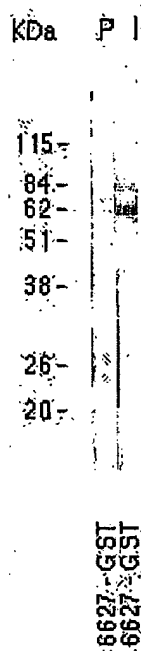
**FIG. 117A**



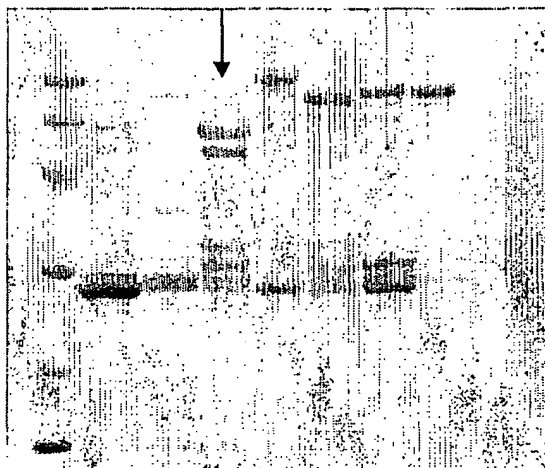
**FIG. 117B**



119/169

**FIGURE 120****Fig. 120A****Fig. 120B**

118/169

**FIGURE 119****FIG. 119A**

kDa P I

115 -  
84 -  
62 -  
51 -  
38 -  
26 -  
20 -

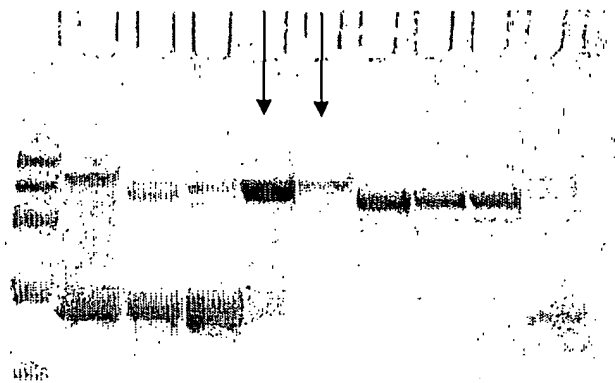
**FIG. 119B**

6528 - GST  
6528 - GST

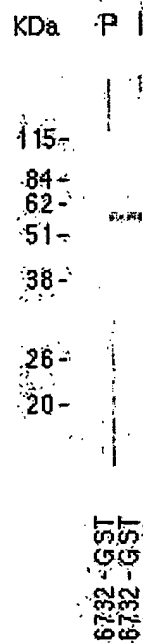
121/169

**FIGURE 122**

**Fig. 122A**



**Fig. 122B**

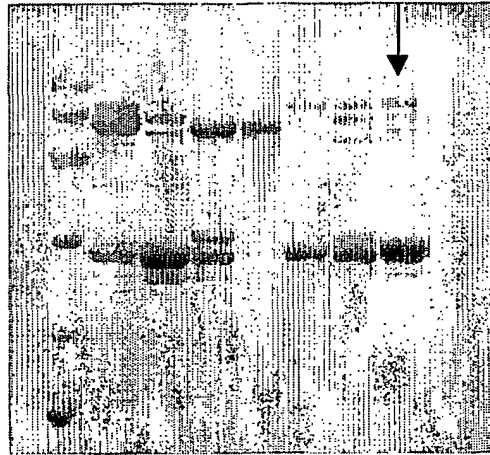




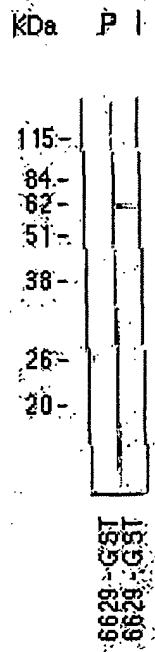
120/169

**FIGURE 121**

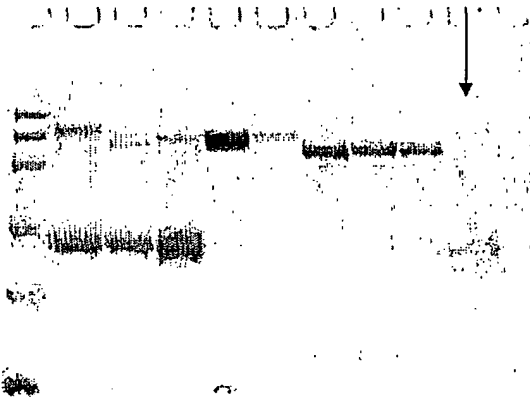
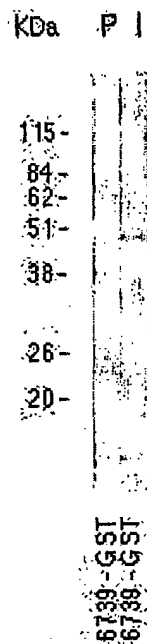
**FIG. 121A**



**FIG. 121B**

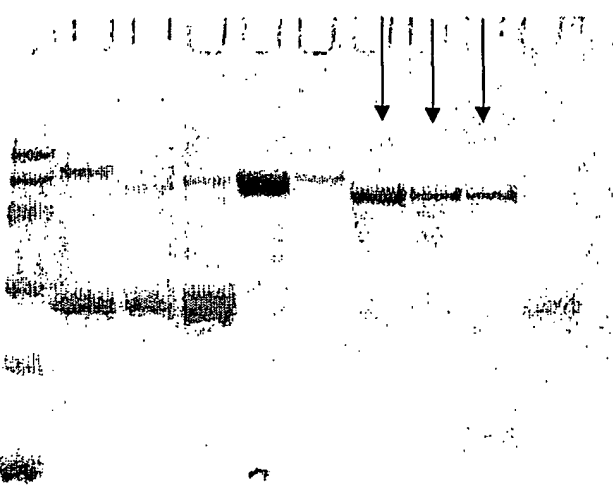


123/169

**FIGURE 124****FIG. 124A****FIG. 124B**

122/169

**FIGURE 123**



**Fig. 123A**

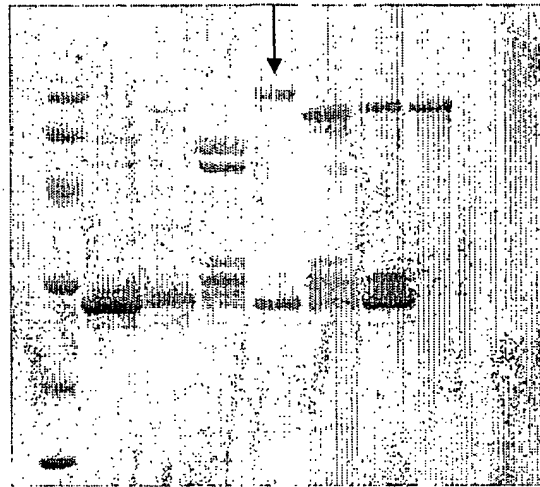
kDa P I

115-  
84-  
62-  
51-  
38-  
26-  
20-

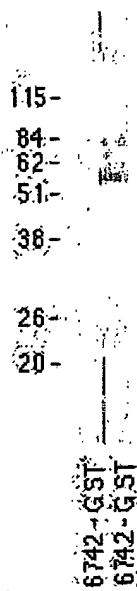
**Fig. 123B**

6738-GST  
6738-GST

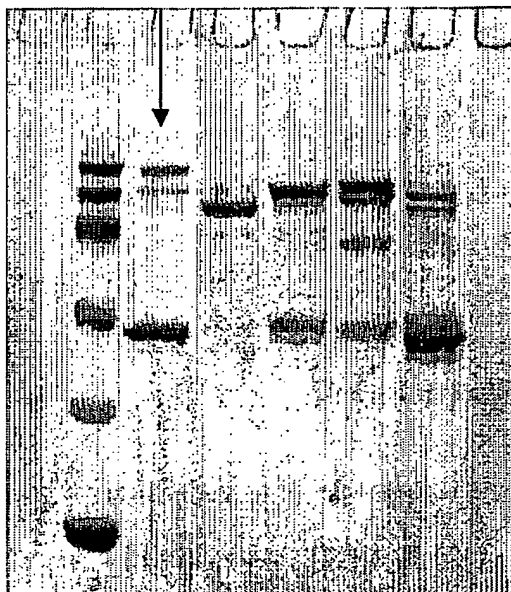
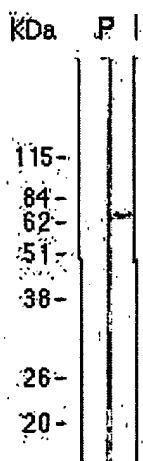
125/169

**FIGURE 126****Fig. 126A**

kDa P I

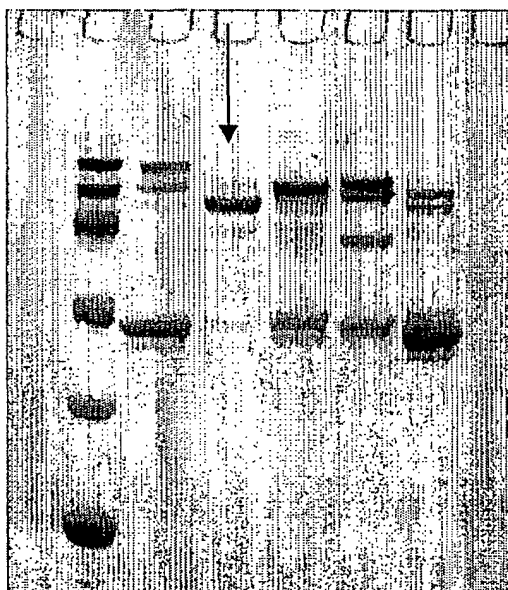
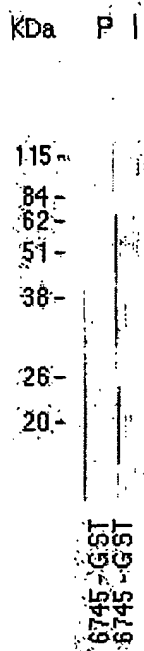
**Fig. 126B**

124/169

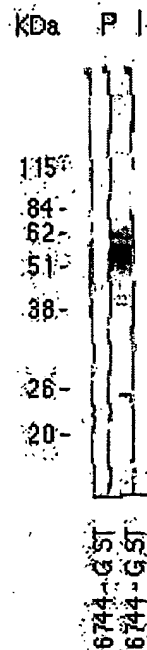
**FIGURE 125****FIG. 125A****FIG. 125B**

6741-GST  
8741-GST

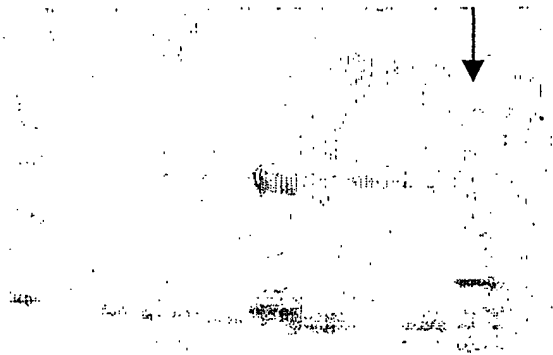
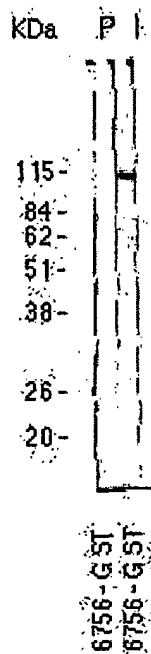
127/169

**FIGURE 128****Fig. 128A****Fig. 128B**

126/169

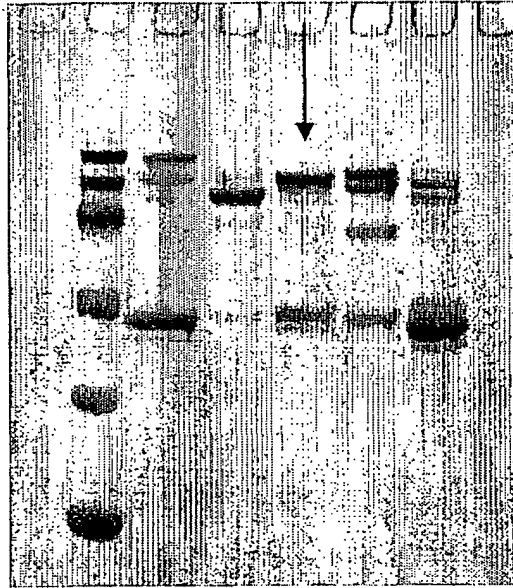
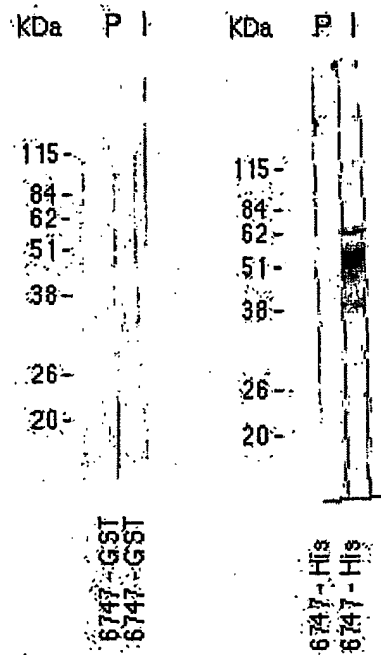
**FIGURE 127****FIG. 127A****FIG. 127B**

129/169

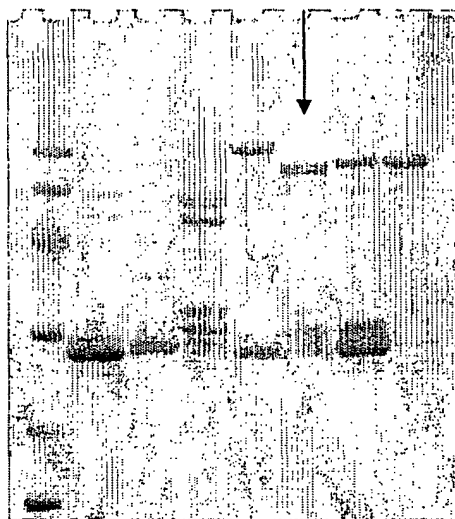
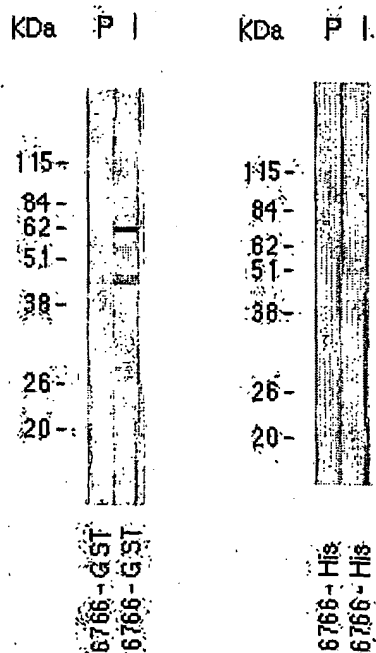
**FIGURE 130****Fig. 130A****Fig. 130B**



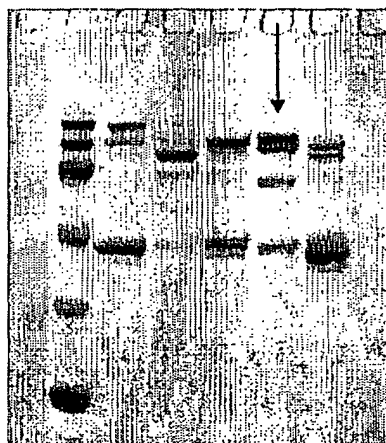
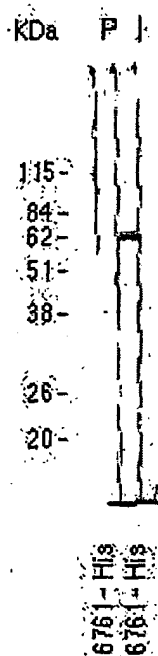
128/169

**FIGURE 129****FIG. 129A****FIG. 129B**

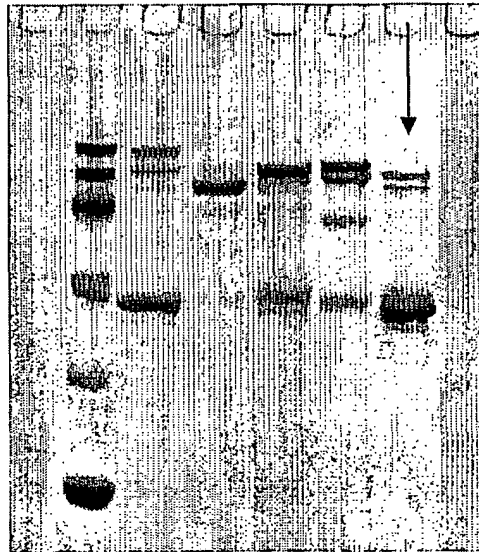
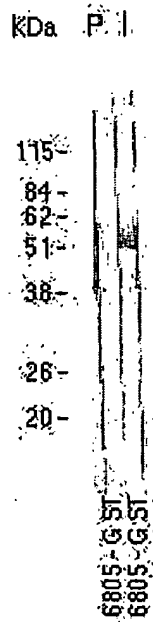
131/169

**FIGURE 132****Fig. 132A****Fig. 132B**

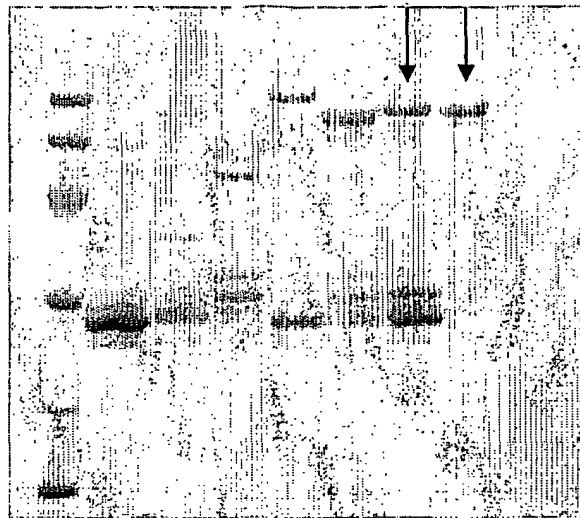
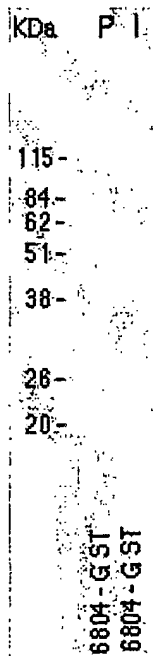
130/169

**FIGURE 131****FIG. 131A****FIG. 131B**

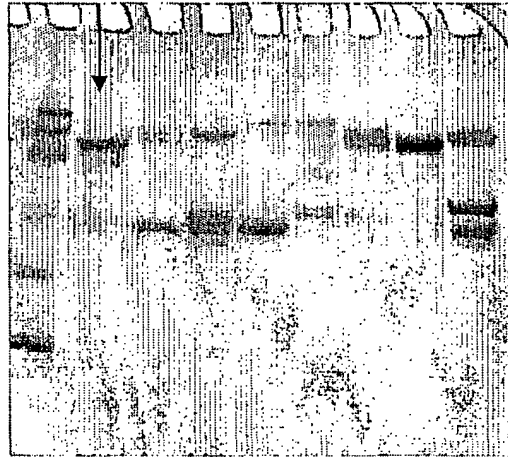
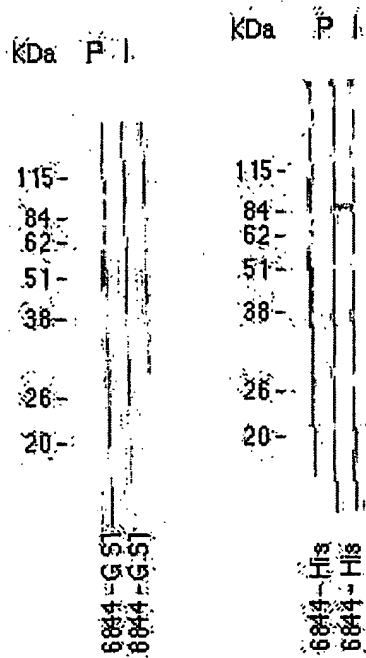
133/169

**FIGURE 134****Fig. 134A****Fig. 134B**

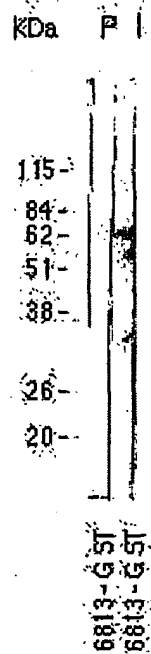
132/169

**FIGURE 133****Fig. 133A****Fig. 133B**

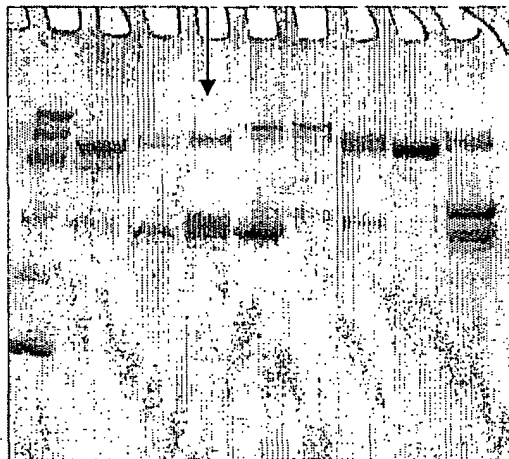
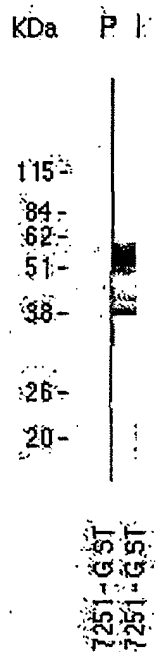
135/169

**FIGURE 136****Fig. 136A****Fig. 136B**

134/169

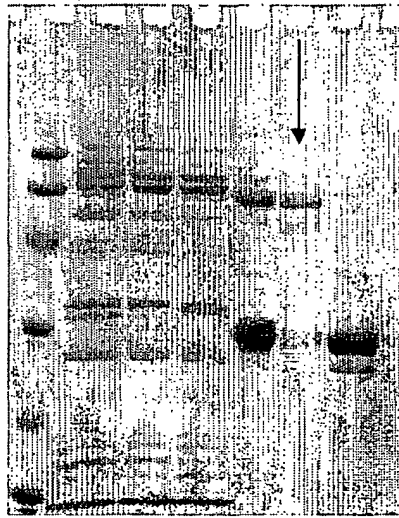
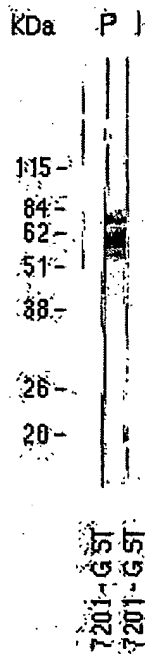
**FIGURE 135****Fig. 135A****Fig. 135B**

137/169

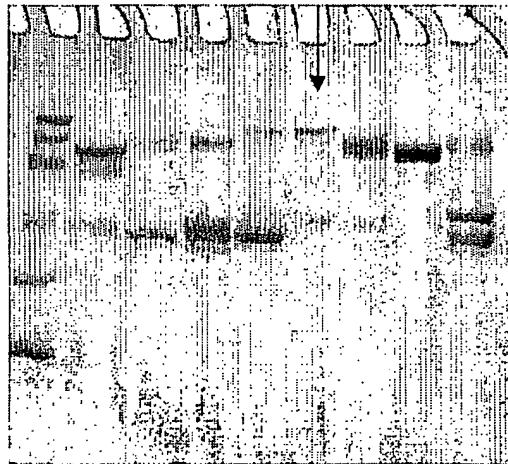
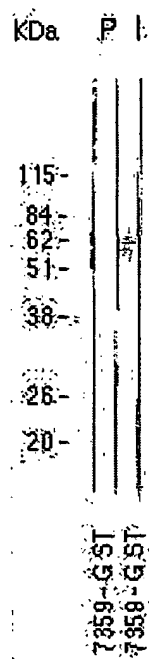
**FIGURE 138****Fig. 138A****Fig. 138B**



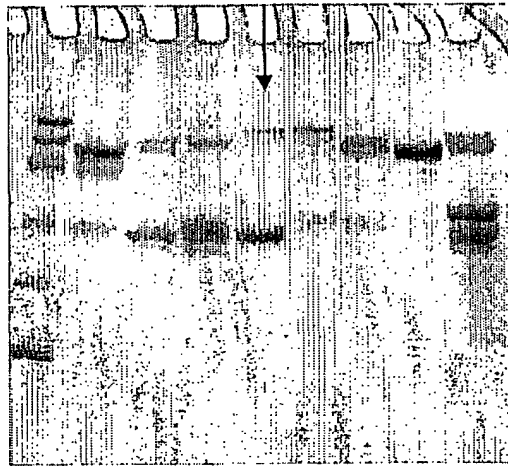
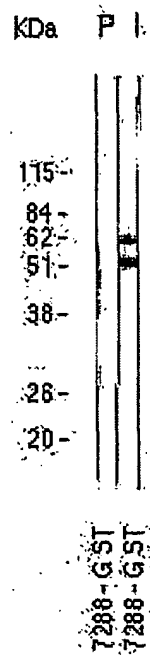
136/169

**FIGURE 137****FIG. 137A****FIG. 137B**

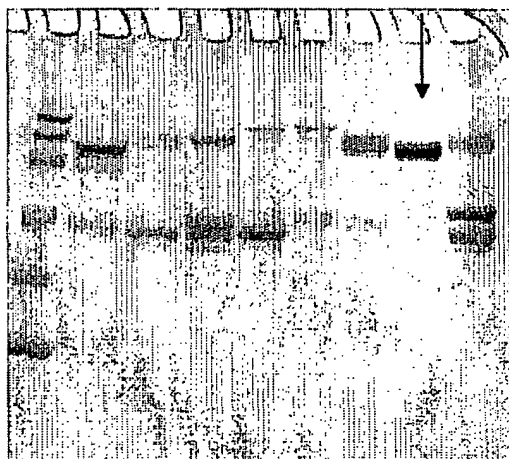
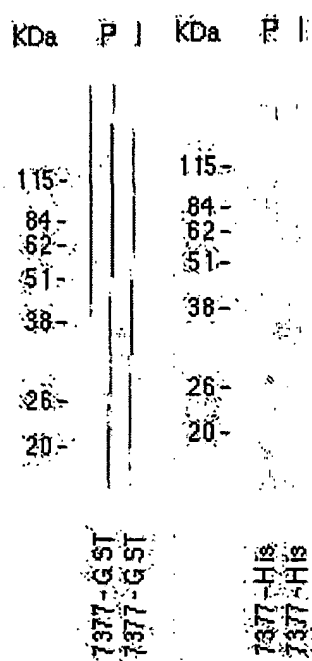
139/169

**FIGURE 140****FIG. 140A****FIG. 140B**

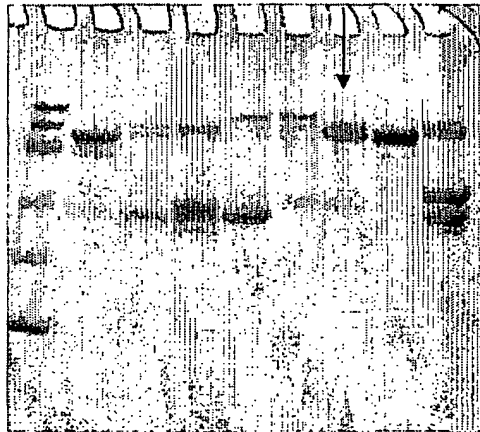
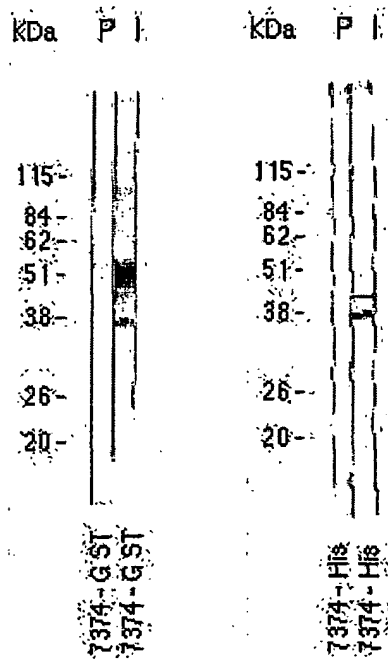
138/169

**FIGURE 139****FIG. 139A****FIG. 139B**

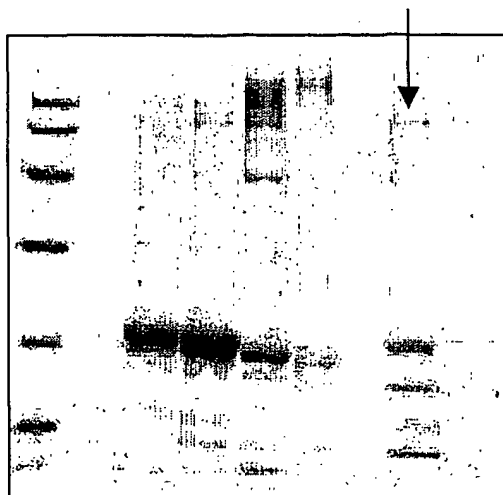
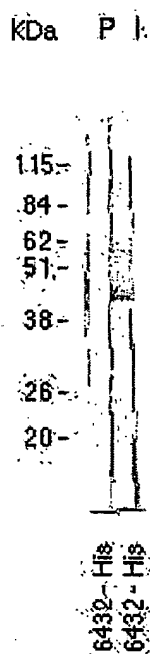
141/169

**FIGURE 142****FIG. 142A****FIG. 142B**

140/169

**FIGURE 141****FIG. 141A****FIG. 141B**

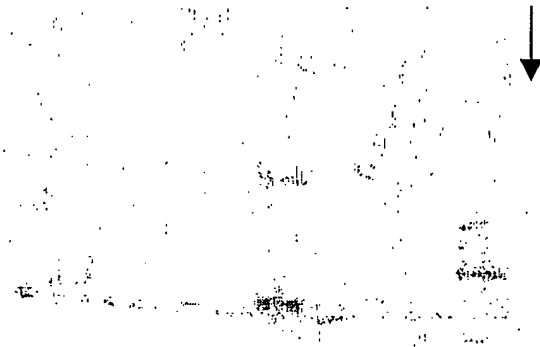
143/169

**FIGURE 144****FIG. 144A****FIG. 144B**

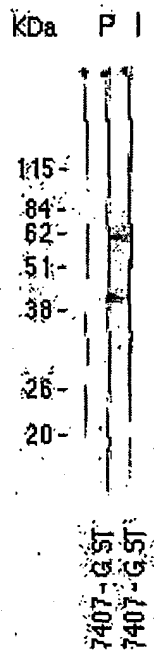
142/169

**FIGURE 143**

**FIG. 143A**



**FIG. 143B**



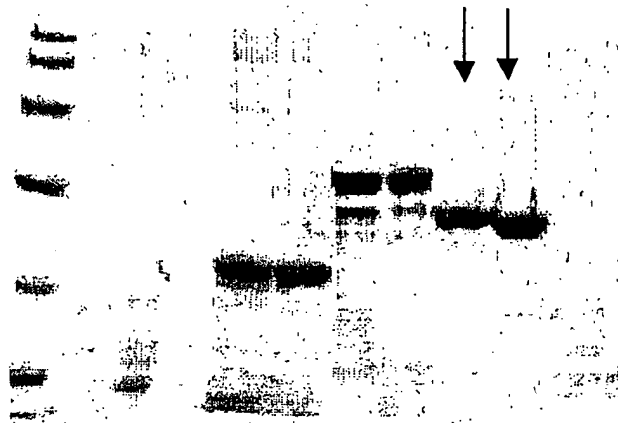




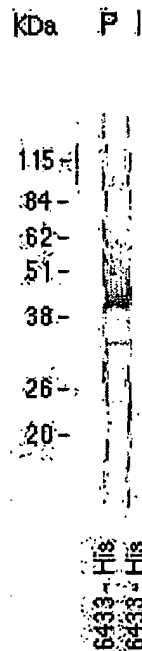
144/169

**FIGURE 145**

**FIG. 145A**



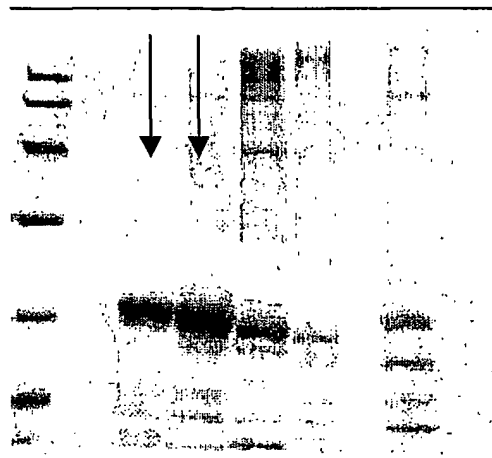
**FIG. 145B**



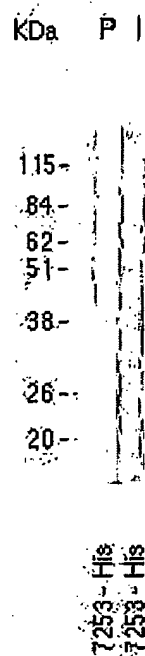
147/169

**FIGURE 148**

**Fig. 148A**



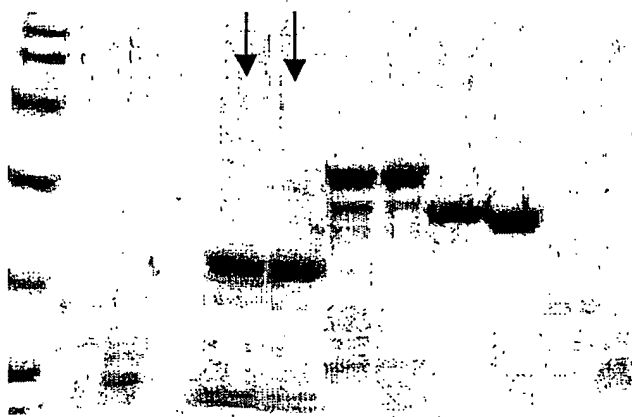
**Fig. 148B**



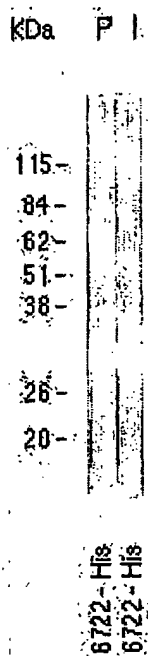
146/169

**FIGURE 147**

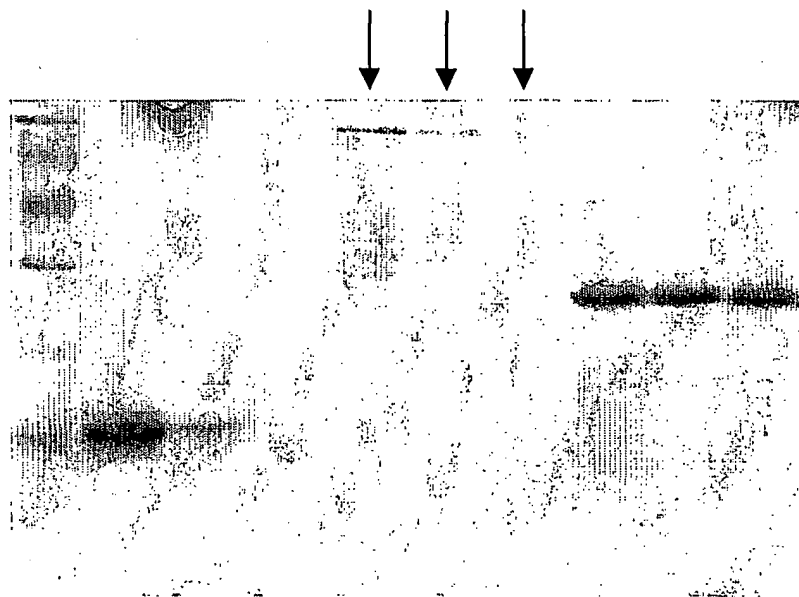
**FIG. 147A**



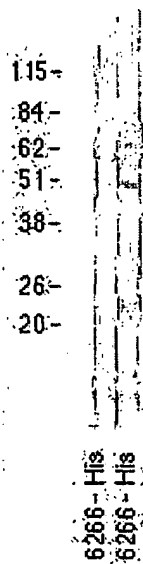
**FIG. 147B**



149/169

**FIGURE 150****Fig. 150A****Fig. 150B**

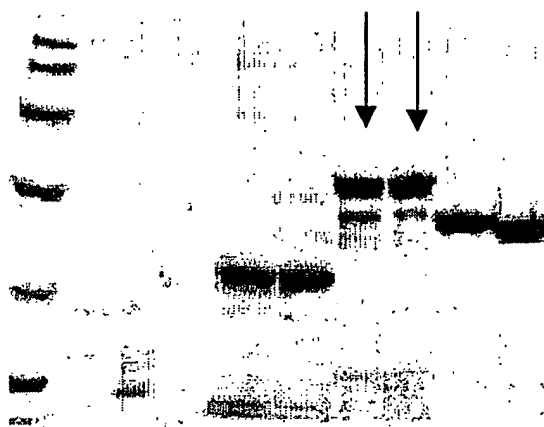
kDa P I.



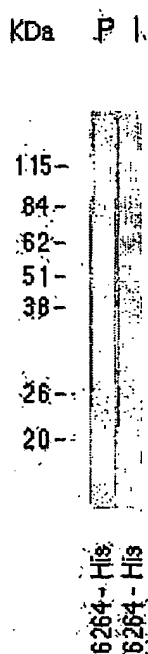
148/169

**FIGURE 149**

**FIG. 149A**



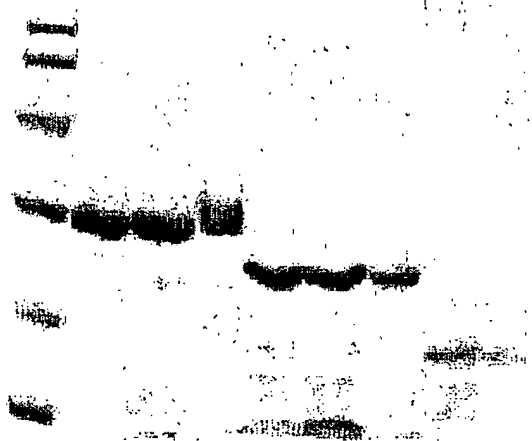
**FIG. 149B**



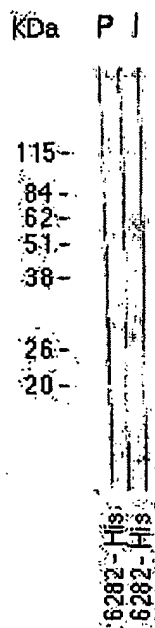
151/169

**FIGURE 152**

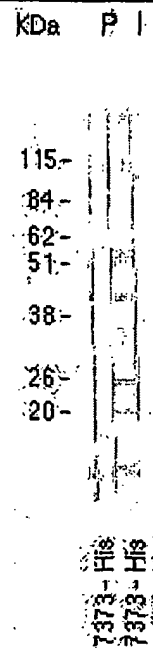
**Fig. 152A**



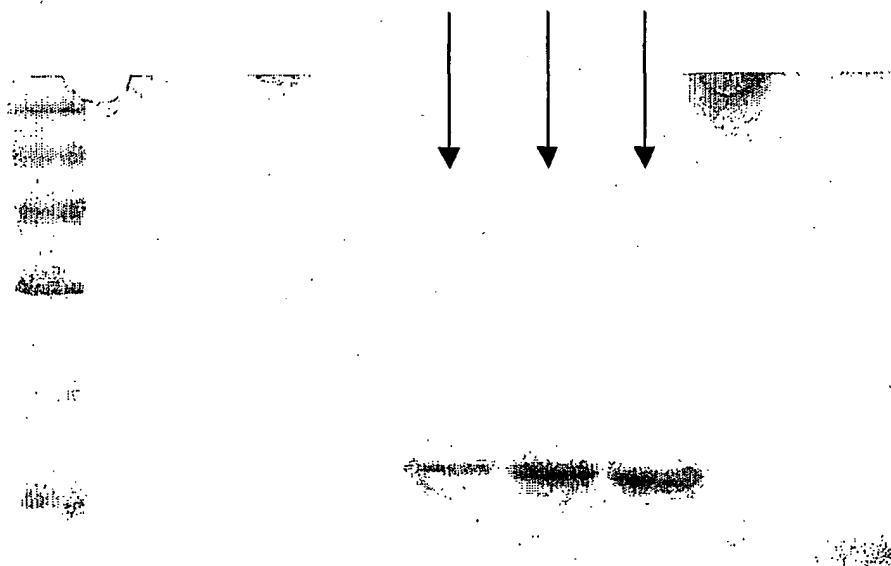
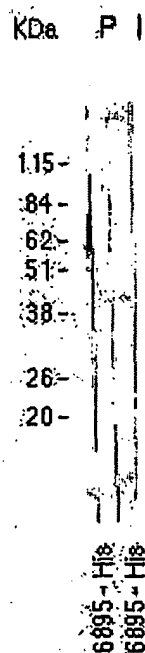
**Fig. 152B**



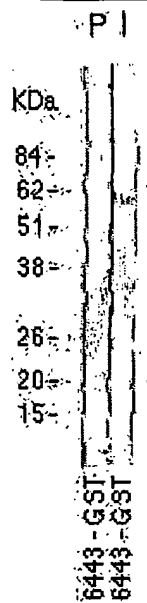
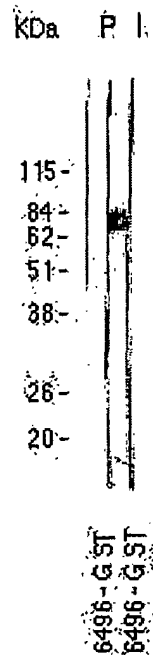
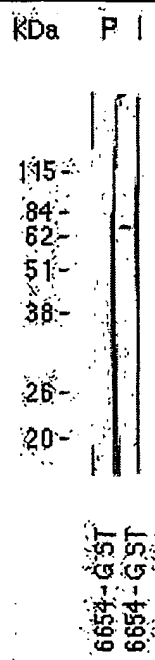
**FIGURE 153**



150/169

**FIGURE 151****Fig. 151A****FIG. 151B**

153/169

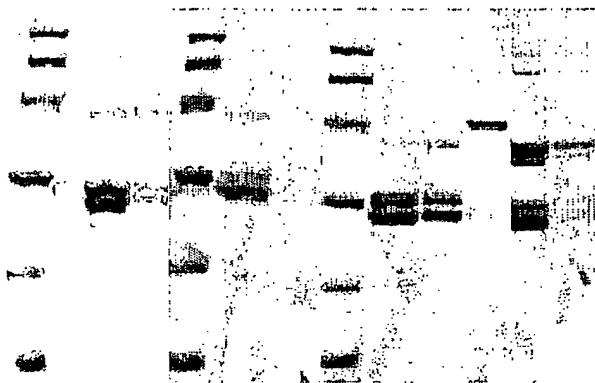
**FIGURE 156****FIGURE 157****FIGURE 158**



152/169

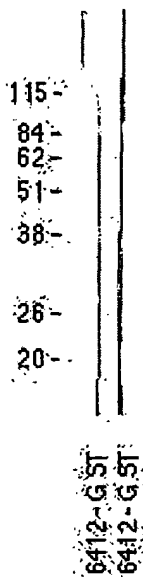
**FIGURE 154**

**Fig. 154A**



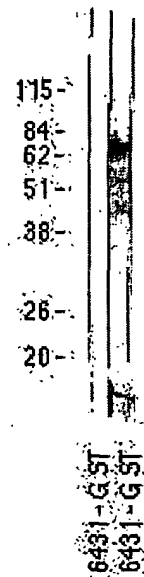
**Fig. 154B**

kDa P I



**FIGURE 155**

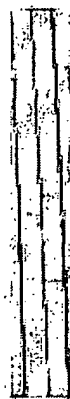
kDa P I



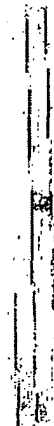
155/169

**FIGURE 161****FIG. 161A****FIG. 161B**

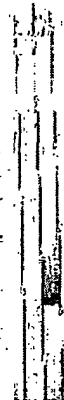
kDa P |

115-  
84-  
62-  
51-  
38-  
26-  
20-644.1-His  
644.1-His**FIGURE 162**

kDa P |

115-  
84-  
62-  
51-  
38-  
26-  
20-His  
His  
674.8  
674.8**FIGURE 163**

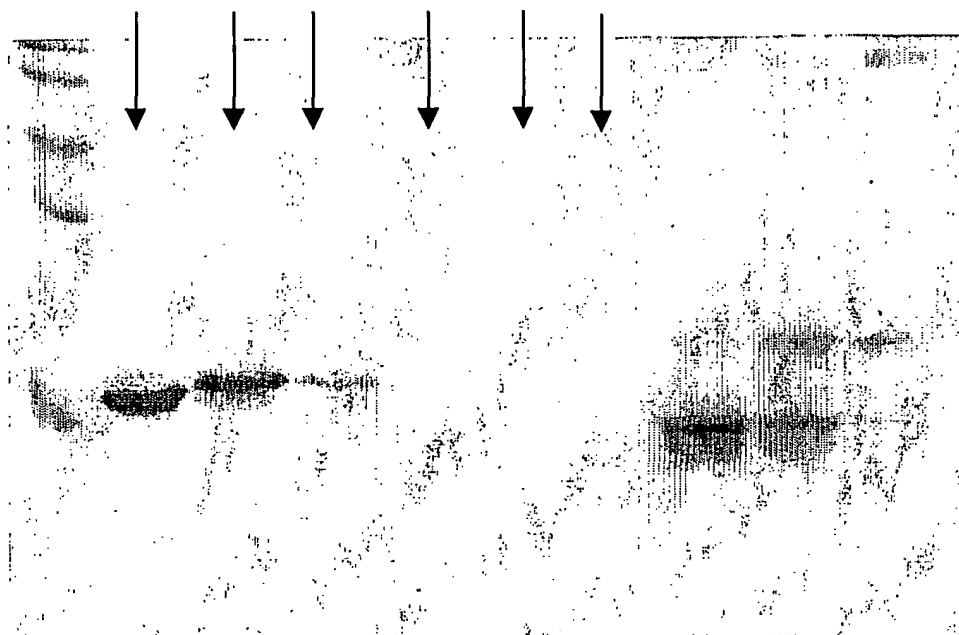
kDa P |

115-  
84-  
62-  
51-  
38-  
26-  
20-His  
His  
688.1  
688.1

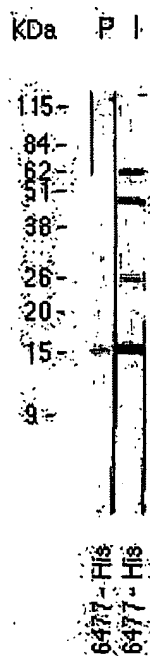
154/169

**FIGURE 159**

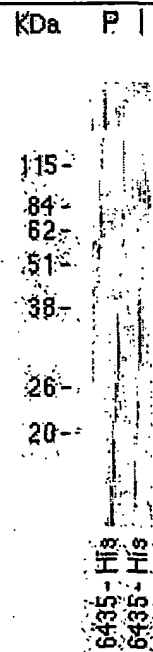
**Fig. 159A**



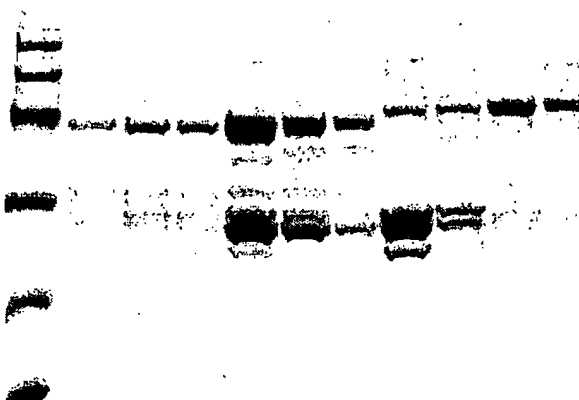
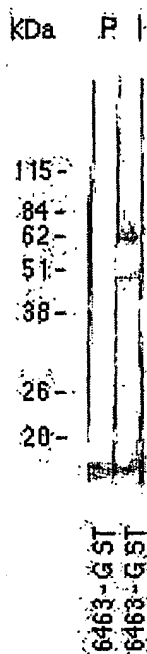
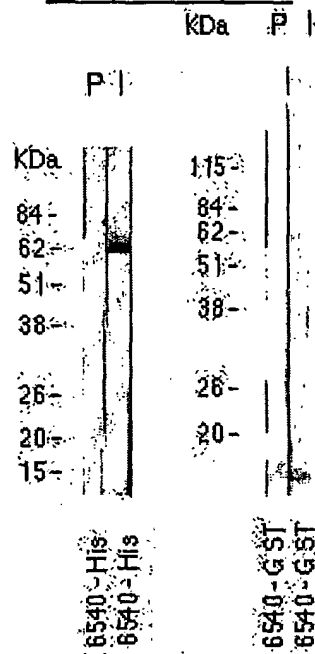
**Fig. 159B**



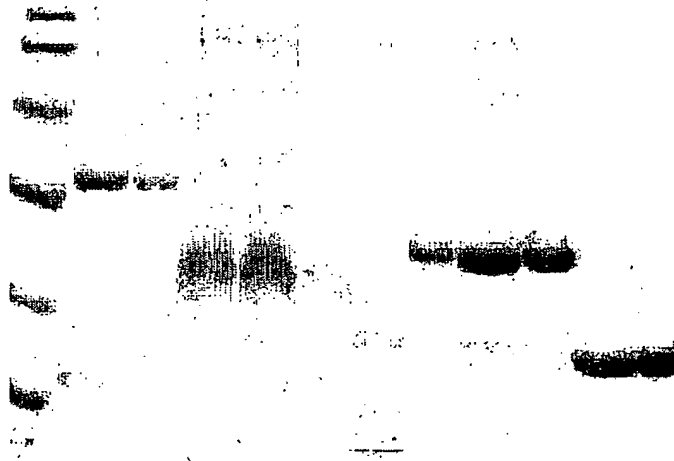
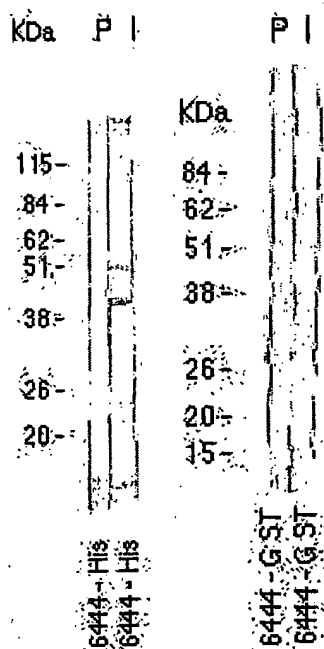
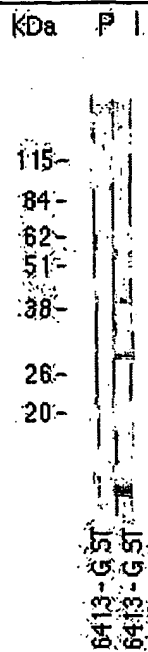
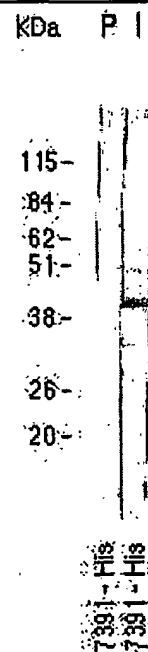
**FIGURE 160**



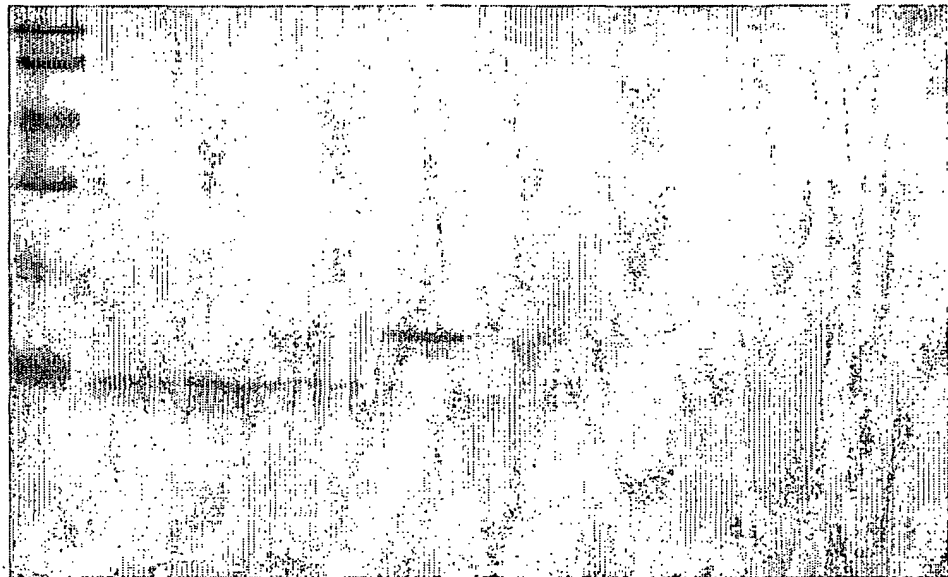
157/169

**FIGURE 167****Fig. 167A****FIG. 167B****FIGURE 168**

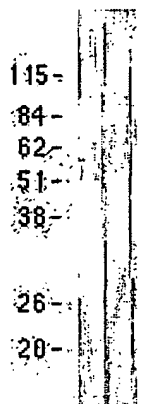
156/169

**FIGURE 164****Fig. 164A****Fig. 164B****FIGURE 165****FIGURE 166**

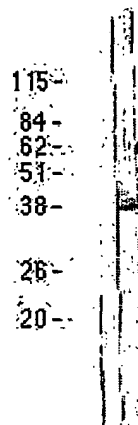
159/169

**FIGURE 171****Fig. 171A****Fig. 171B**

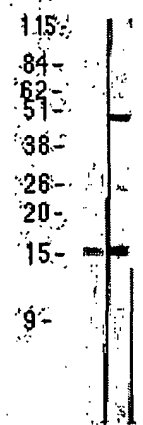
kDa P I

6632 - His  
6632 - His**FIGURE 172**

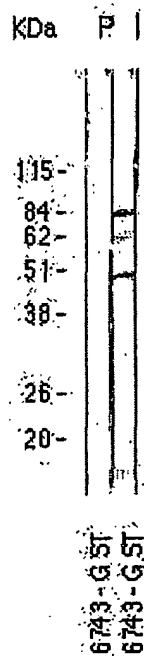
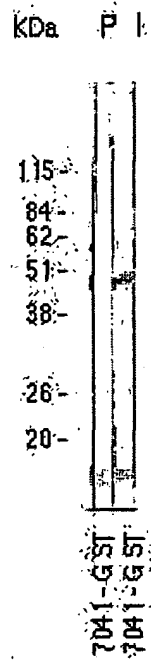
kDa P I

6748 - His  
6748 - His**FIGURE 173**

kDa P I

6497 - His  
6497 - His

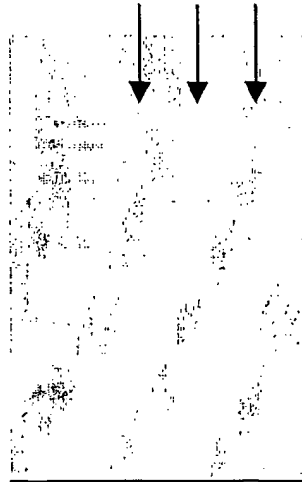
158/169

**FIGURE 169****FIGURE 170**

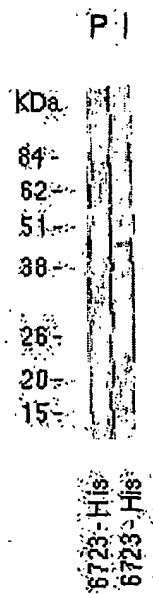
161/169

**FIGURE 179**

**Fig. 179A**

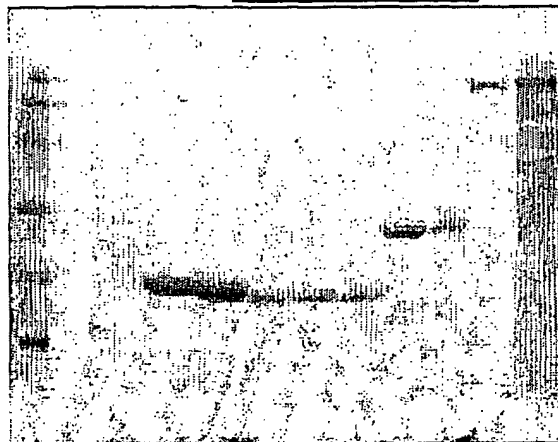
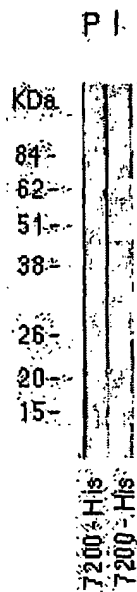
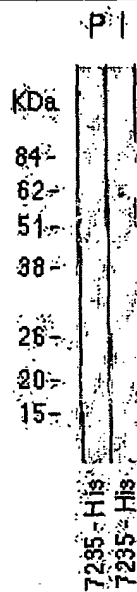
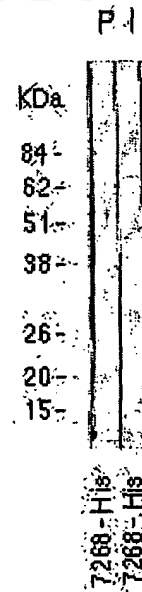
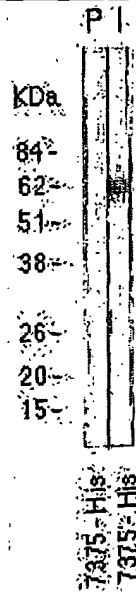
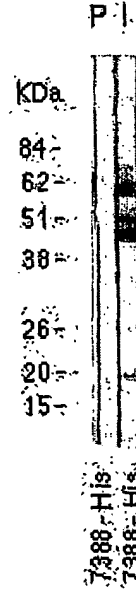


**Fig. 179B**





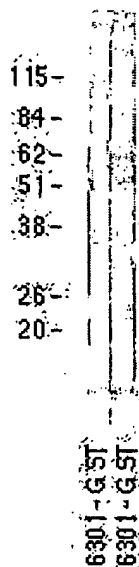
160/169

**FIGURE 174****Fig. 174A****Fig. 174B****FIGURE 175****FIGURE 176****FIGURE 177****FIGURE 178**

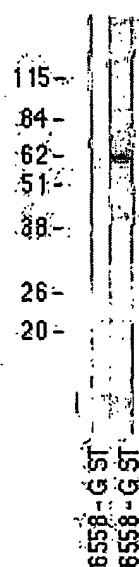
163/169

**FIGURE 181**

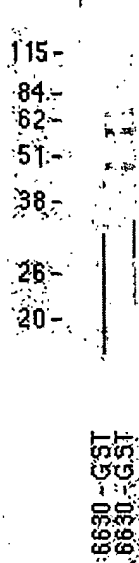
KDa P I

**FIGURE 182**

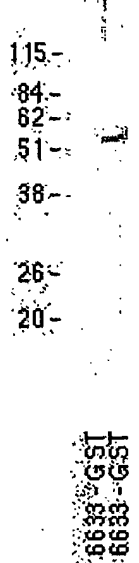
KDa P I

**FIGURE 183**

KDa P I

**FIGURE 184**

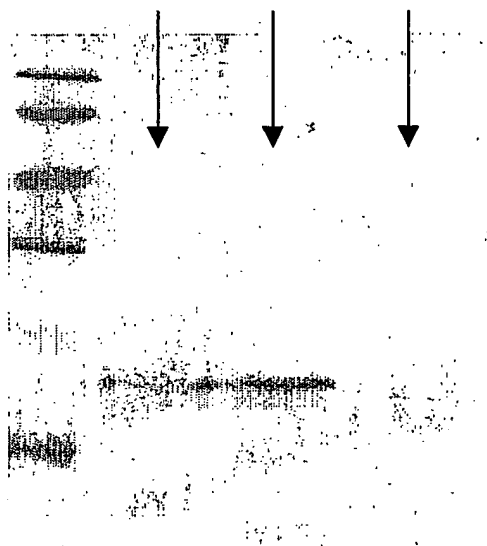
KDa P I



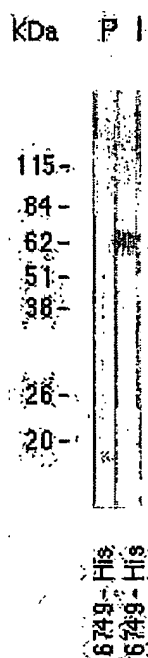
162/169

**FIGURE 180**

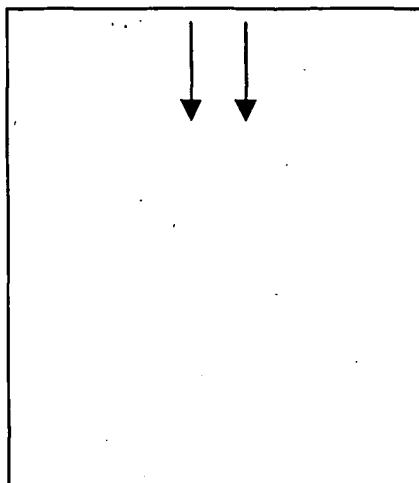
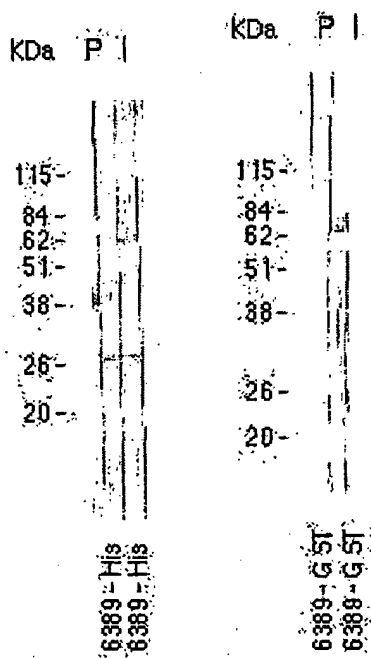
**Fig. 180A**



**Fig. 180B**



165/169

**FIGURE 186****Fig. 186A****Fig. 186B**

164/169

**FIGURE 185**

kDa P I

115-

84-

62-

51-

38-

26-

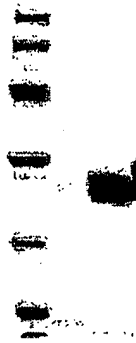
20-

6642-GST  
6642-GST

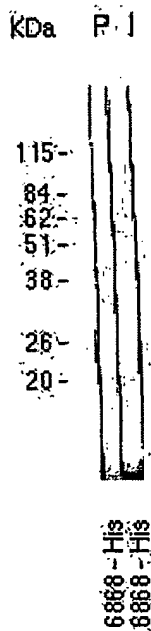
167/169

**FIGURE 188**

**Fig. 188A**



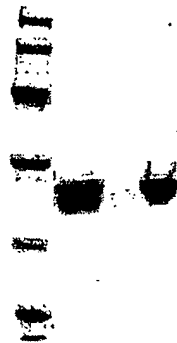
**Fig. 188B**



166/169

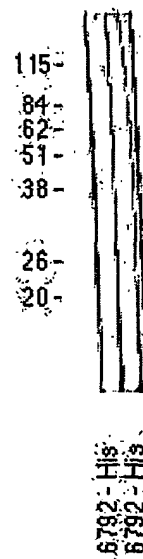
**FIGURE 187**

**Fig. 187A**

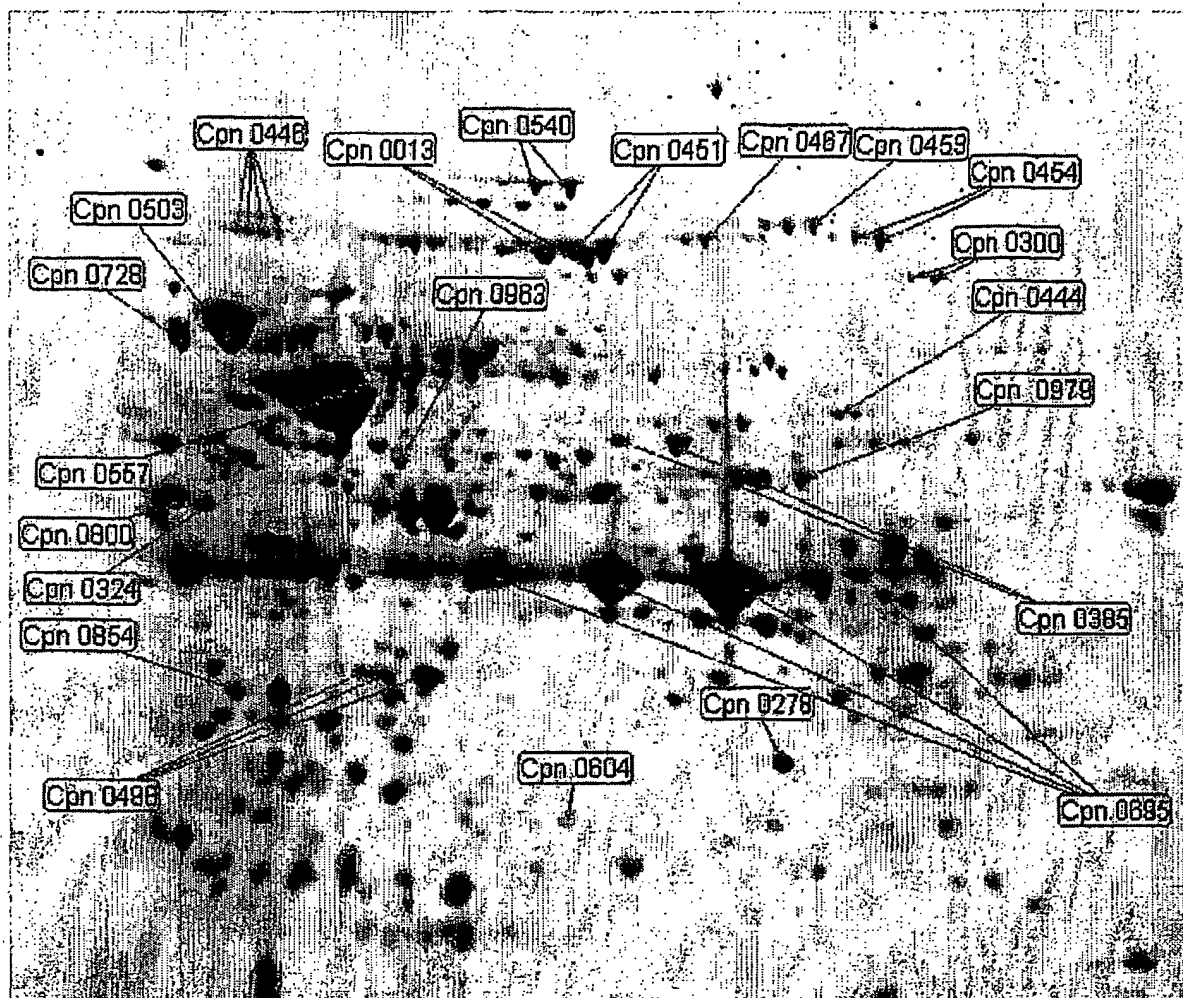


kDa P. I.

**Fig. 187B**

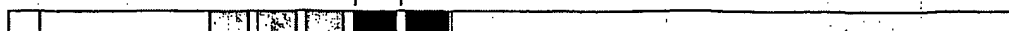


169/169

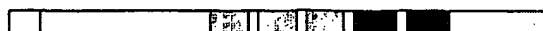
**FIGURE 190****FIGURE 191**

SVIVG.VSTNSEHRYHAFQYADGQMVDLGTLCGPESYAQGVSGDGK  
 KVIIVG.HSTRIDGEYRAFKYVDGRMIDLGTLCGSASFAGVSDDGK  
 KVIIVG.RSETYYGEVHAFCHKNGVMSDLGTLCGSYSAAKGVSATGK  
 KVIIVG.WSTTNNGETHAFMHKDETMHDLGTLCGGFSVATGV SADGR  
 TIIVGSMESTITRKTTAVKWVNNVPTYLGTLCGDASTGLYISGDGT

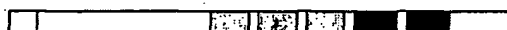
7107



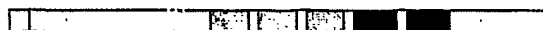
7109



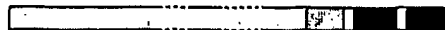
7110



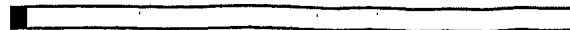
7108



7105



7106

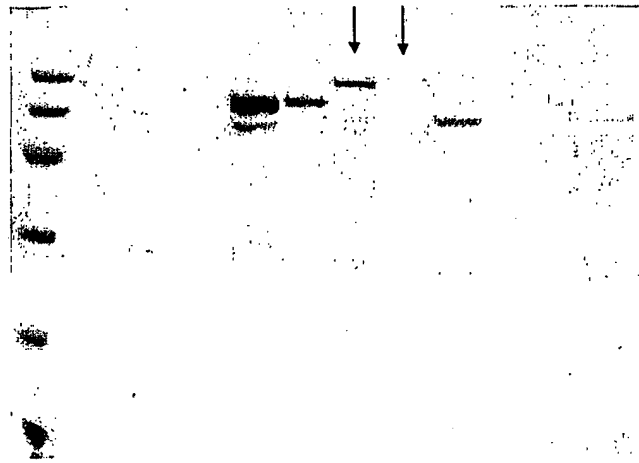




168/169

**FIGURE 189**

**FIG. 189A**



**FIG. 189B**

